DISCOVERING THE LANGUAGE OF MEANINGFUL WORK

By

Michael Aubrey Morrison

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology — Master of Arts

2018

ABSTRACT

DISCOVERING THE LANGUAGE OF MEANINGFUL WORK

By

Michael Morrison

This study introduces a series of language signals that indicate whether a person finds their work meaningful (or meaningless). These signals are then integrated into a new, natural language measure of work meaningfulness. This algorithm can analyze a worker's written description of their work, and using features of their writing determine whether they find their work meaningful with an average classification accuracy of 85%. As an additional, theoretical contribution, this study tests the relationship between work meaningfulness and construal level theory. Results indicate that personal pronouns and action verbs are most related to creating an impression of meaningfulness, but that identity statements and positive sentiment are more related to actual, self-reported meaningfulness. Additionally, construal level showed a significant, positive relationship with several measures of work meaningfulness.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

"So, what do you do for a living?" we ask at nearly every social event we attend with new people. Sometimes a short answer to this question is sufficient. Other times, when we're interested (or bored), we encourage people to expand further, to tell us *all* about their work. When you listen to these longer answers, it's often easy to get a sense of how your conversation partner feels about their work: Whether they like it, hate it, think of it as temporary, or find it meaningful or meaningless. But what cues in their language lead you to these conclusions?

What if we could analyze and measure those cues in language *directly*? To determine, in particular, whether a person finds their work meaningful just by how they talk about it? The potential upsides to being able to measure work meaningfulness in language extend far beyond a simple party trick. People are constantly talking about their work in contexts where it would be useful to know how much meaning they find in it.

Imagine all the job interviews being conducted right now, where some job candidate is recounting their past jobs. How meaningful did they find those past jobs? Imagine consultants, sitting at conference tables asking employees to "tell me a little bit about what you do here." What verbal cues might alert those consultants to workers with high or low meaningfulness? Finally, there is a wealth of text data being generated every day: emails, internet chats, telephone transcripts, etc. where workers discuss their job roles to leaders, customers, and fellow coworkers.

Natural language descriptions of work that potentially contain rich information about work meaningfulness are happening everywhere all around us, and we're not measuring any of it! The only way we can measure meaningfulness right now is by asking people to bubble-in their agreement on a series of Likert survey items. Putting aside the inconveniences of issuing

1

surveys, what might we be missing about the construct of meaningfulness by only measuring it via traditional Likert surveys? What could we learn about what meaningfulness is — how to define the experience of meaningful work — from measuring how it shows up in people's language?

In this study, I asked **n=194** full-time workers to tell me all about what they do for a living, and then to reveal whether they find their work meaningful. Using Machine Learning (ML), I analyzed each person's 'work story' to discover linguistic signals associated with feelings of high or low meaningfulness. I then integrated these signals into a new algorithm — called a Natural Language Processor (NLP). This NLP algorithm can read-in any work story and output a probabilistic conclusion about whether the author of that text finds their work meaningful.

As a third contribution of this study, I also used language analysis to test the relationship between a sense of meaningfulness in work and a cognitive-psychological construct called construal level, which has the potential to help push meaningful work theory towards a consensus about what meaningfulness *is*.

Meaningfulness: A 'Holy Grail' Variable in Organizations

Work meaningfulness is like a super-variable for organizations. If an organization can lead its members to a strong sense of work meaningfulness, that organization can enjoy workers who are more committed, more motivated, who go above-and-beyond, can manage their stress, and perform well at work (Bunderson & Thompson, 2009; Champoux, 1992; Glazer, Kozusznik, Meyers, & Ganai, 2014; Grant, 2008; Seibert, Wang, & Courtright , 2011). Meaningfulness, in effect, leads to nearly all the positive employee outcomes that organizations care about achieving.

Understandably, there is a great deal of research dedicated to figuring out how to foster a sense of meaningfulness at work. And these efforts have found that meaningfulness is related to some of the most central concepts in organizational science, including: job design, transformational leadership, work engagement, and job fit (Arnold et al., 2007; Brit, Adler, & Bartone, 2001; Hackman & Oldham, 1976; May, Gilson, & Harter, 2004; Rothmann & Hamukangandu, 2013). Put simply, meaningfulness is a very *desirable* phenomenon in organizations, and that is reflected in its frequent use as a criterion of interest in the organizational science literature.

While much of the literature in organizational science seems to focus on the causes and consequences of meaningful work for organizations, usually as a secondary focus to some other construct (e.g., meaningfulness as an outcome in a study about job design), more focused articles on meaningfulness can be found in the wider psychology literature that extends to other sub-disciplines such as social and vocational psychology. This meaning-centric literature focuses primarily on defining exactly *what meaningfulness is* and how it is constructed from a person-centric perspective (Weiss & Rupp, 2011).

What Does 'Meaningfulness' Mean?

When a person says that they find their work 'meaningful' or 'meaningless', what do they mean? Philosophers and psychologists have been trying to capture and define the notion of meaning in work for some time. In a recent review, Bailey, Yeomen, Madden, Thompson, and Kerridge (2016) listed over 30 different definitions of the term "work meaningfulness." And although the last decade has seen some popular, highly-integrative reviews of the meaningfulness space (e.g., Bailey et al., 2016; Lepisto and Pratt, 2016; Rosso et al., 2010) which have included their own 'universal' definitions of meaningfulness, there is still no

dominant, generally-accepted definition or model of work meaningfulness.

Common themes. While the construct proliferation currently challenging the work meaningfulness literature may hinder consensus, it is a great boon to comprehensiveness --- as in, towards identifying all the potential factors in meaningfulness, and potential routes to achieving it. And while there is not yet a consensus on what work meaningfulness is or what creates a sense of meaningfulness in work, there are certainly some common themes that are mentioned often when attempting to define, discuss or test experiences of meaningfulness. In the following sections, I will list these common themes. For each theme, I will provide a theoretical background and an example of a definition of work meaningfulness that relies on the theme.

Higher purpose. Seminal psychologists Victor Frankl (1962) and Abraham Maslow (1943) both emphasized the notion of self-transcendence as a path to creating meaning in life. Although there are a few different conceptualizations of self-transcendence in the literature, the essential idea is that self-transcendence is the experience of seeing your actions as serving a goal that is beyond yourself (Morrison, 2016). Or as legendary psychologist William James (1985, p. 266) put it, "a feeling of being wider in life than the world's selfish little interests."

The notion of pursuing a higher, self-transcendent purpose was central to Frankl (1962) and Maslow's (1969) definitions of meaningfulness (note that just before he died suddenly while jogging, Maslow published a revised, six-level version of his famous Hierarchy of Needs that included self-transcendence above self-actualization). Self-transcendence is also reflected in modern definitions of meaningful work (Bailey et al., 2016; Rosso et al., 2010): For example. Arnold et al. (2007, p. 195) define meaningful work as "finding a purpose in work that is greater than the extrinsic outcomes of the work."

Prosocial impact. Although Frankl (1962) and Maslow's (1943) theoretical arguments do

4

not mandate that the self-transcendent 'higher purpose' of your work needs to be 'helping other people', that theme (helping others, or 'prosocial impact' in modern terms) is featured frequently in their writings and in the examples they employed to illustrate what 'finding a higher purpose' might look like.

In modern meaningfulness literature, the theme of prosocial impact features prominently. Hackman and Oldham (1976, p. 257) proposed *task significance* ("the degree to which the job has a substantial impact on the lives or work of other people") as one of the three job design features that lead to experiencing work as meaningful. Duffy, Allan, Autin, and Bott (2013) included "serving others in some capacity" in their definition of meaningful work. By way of testing these propositions, Allan, Duffy, and Collisson's (2018) study concluded that feeling like you're improving the lives of others with your work leads to a sense of meaningfulness. Further, Grant's (2008) study showed that adding a sense of prosocial impact to a job can increase performance by increasing a sense of task significance.

Return on investment. Kahn (1990, p. 704) defined meaningfulness as "feeling that one is receiving a return on investments of one's self." We need to see our time spent working as fruitful, either simply or profoundly. The notion of needing a return on investment from work is particularly reflected in theories and studies of meaning*less* work. Ariely, Kamenica, and Prelec (2008) argued that the "Sisyphus experience" — referring to the mythical Greek figure who was forced to roll a boulder up a hill only to see it roll back down again (for all eternity) — is a common route to draining meaning from work. Ariely et al. (2008) illustrated this phenomenon in an experiment where participants were instructed to build figures out of LEGO bricks. For the low-meaning condition, experimenters immediately picked up each LEGO figure upon completion, disassembled the figure, and handed the bricks back to the participant. Participants

in this disassembly condition experienced markedly lower meaningfulness than those in the 'high meaning' condition where completed figures were placed on display.

As a final example, Hackman and Oldham's (1976) notion of *task identity* also echoes the 'return on investment' theme. They proposed that being able to see a completed whole produced by one's work is a key component in the experienced meaningfulness of work.

Self-growth. The idea of meaningful self-growth is perhaps best captured by Maslow's (1943) notion of self-actualization. Maslow (1943) placed the achievement of one's full potential at the top of his original *Hierarchy of Needs*, arguing that self-actualizing was the highest aim in life. Self-growth is also discussed frequently in modern meaningfulness literature. For example, Rosso et. al (2010) and Lips-Weirsma and Wright (2012) both feature analogs of self-growth in their models of meaningful work. And Fairlie (2011) defines work meaningfulness explicitly as "work that facilitates self-actualizing."

Self-expression. Self-expression is typically discussed in terms of opportunities that an organization provides for its members to bring their 'whole self' to work (Chalofsky, 2010). For example, a nurse that can express her passion for painting by decorating patients' rooms may find her work more meaningful than if this expression was prohibited or blocked. Many modern theories of meaningful work support the notion that self-expression is related to meaningfulness: Hackman and Oldham (1976) listed *skill variety* (the degree to which a job "involves the use of a number of different skills and talents of the person") as one of their three pillars of work meaningfulness. Finally, major integrative reviews of the meaningfulness space by Rosso et al. (2010) and Lepisto and Pratt (2016) both included a form of self-expression in their components of meaningfulness.

Sensemaking. In the meaningfulness literature, the words 'meaning' and

'meaningfulness' are highly distinct (Rosso et al., 2010). Work *meaning* is what work represents or symbolizes. It is how you 'justify' what you're doing in your work (Lepisto & Pratt, 2016). Through this lens, meaning is the outcome of the process of making sense of your job and work role (Martela & Steger, 2016; Rosso et al., 2010). The meaning you learn from your work is thought to spill-over into the rest of your life, helping you understand the world and your place in it. Schnell, Hoge, and Pollet (2013) defined work meaningfulness in terms of this kind of sensemaking — as your work providing you with a broader understanding. Steger et al., (2012) also featured a sensemaking-themed item in their highly-cited measure of work meaningfulness (the item text reads "My work helps me make sense of the world around me").

    Work Centrality and Job Involvement. How important is your work in your life? Your work centrality is determined by how important you find your work, relative to other aspects of your life (Rosso et al., 2010). Workers who get a lot of meaning from their work often see their work as an extremely important part of their lives (i.e., have high work centrality; Bunderson & Thompson, 2009). Similarly, job involvement is a measure of how 'wrapped up' your sense of self is in your work. While having high work centrality and/or job involvement is thought to be associated with a greater sense of meaningfulness in work, work centrality and job involvement can also 'cut both ways.' That is, highly involved workers are more emotionally sensitive to developments in their work life, both positive *and negative* (Douglas & Carless, 2009).

    Identity. More than work being important to you, work can define you and help you understand who you are, perhaps especially when it is meaningful. Some scholars have suggested that highly meaningful work is work where an individual "connects [their] identity to his or her work" (Britt et al. [2007] p36) or "integrates their personal identity with their work role" (Cohen-Meitar, Carmeli, & Waldman, 2009). Rosso et al.'s (2010) integrative review also

features "identity affirmation" as a pathway to meaningfulness, and they noted that identity development seemed to be the most prominent mechanism of meaningfulness presented in research focusing on the organization as a source of meaning (as in, organizations can foster meaningfulness by helping workers find a sense of identity through their work).

Universal definitions. Ultimately, we need to arrive at a single, unifying definition of what meaningfulness is. Chiefly, we need a consensus definition so that we can test and measure meaningfulness in a consistent fashion. Right now, there are so many different definitions and measures of meaningfulness that the few empirical studies that have been conducted on meaningfulness often define, measure, and operationalize meaningfulness completely differently from each other (Bailey et al., 2016). This limits our ability to build a deep understanding of meaningfulness by discovering all the instantiations, contexts, effects, and boundary conditions that stem from a single theoretical root.

Meaningful work researchers are well aware of this need for a universal definition, and there have been many attempts to create one.

Combinatorial definitions. Combinatorial definitions attempt to create a unified definition of meaningfulness by incorporating several of the most common themes. That is, they define meaningfulness as the occurrence of one or more specific meaningfulness themes (Bailey et al., 2016). For example, McCrea, Boreham, and Ferguson (2011) defined meaningfulness as "perceived creativity, autonomy, responsibility and contribution to society."

The trouble with combinatorial definitions such as this one is that none of them cover the full variety of the identified themes commonly associated with meaningful work. Rosso et al. (2010) alone identified 13 different pathways to meaningfulness. To comprehensively define meaningfulness in terms of all the narrow pathways *to* meaningfulness, it would likely take a

long definition.

Broad language definitions. In an attempt to succinctly define meaningful work in a universal manner that accounts for all identified *and yet-to-be-discovered* pathways to meaningfulness, 'broad language' definitions of work meaningfulness employ more abstract language and umbrella terms that, semantically, can accommodate any specific instantiation of meaningfulness within them. This includes defining meaningful work as "important" or "valuable" and/or "significant" (Bailey et al., 2016; Rosso et al, 2010). For example, Hackman and Oldham (1975) defined meaningfulness as work that is "important, valuable, and worthwhile."

This broad language approach to defining meaningfulness is useful because it helps capture how people feel about their work when it is meaningful, regardless of *why* they feel it is meaningful. For example, your work can feel important and valuable *because* it contributes to your self-growth, or because it has a higher purpose or a prosocial impact. In this way, these succinct, broad language definitions are a step towards a universal definition, but not all the way there yet.

The trouble with many of these broad language definitions is that they often rely on synonyms for meaningfulness. A synonym, even a very good synonym, is not the same thing as a definition. A person's answer to the question "Is your work important?" may be different than their answer to the question "Is your work meaningful?" Try this thought experiment: think of the most meaningful work you can imagine for yourself. Got it? OK, now imagine that instead of doing that thing, you are assigned to ladle soup to orphans. You might agree that your soup-ladling is extremely *important* work, even *significant* work, but it may not feel as *meaningful* to you as that first thing you imagined.

9

Unifying with Construal Level Theory. Morrison, Walker, and DeShon (2016) took a different approach to create an all-inclusive, unifying definition of meaningfulness. Rather than seeking to achieve all-inclusiveness and conciseness through semantic manipulations alone, they attempted to define meaningfulness through a common *mechanism* in the experience of work meaningfulness (instead of common descriptors). Drawing from cognitive psychological theory, they proposed that Trope and Liberman's (2010) Construal Level Theory (CLT; detailed below) could be a common cognitive-psychological mechanism underlying *all* experiences of meaningful work. And that this could be used to create a universal definition of meaningfulness that was both as precise as the combinatorial definitions and as comprehensive and concise as the broad language definitions.

Construal Level Theory. Briefly, Construal Level is a sensemaking mechanism through which people think about any given experience in terms of its abstract or concrete qualities (Trope & Liberman, 2010). Put another way, your current construal level is whether you are seeing the forest or the trees. When you are thinking about something at a high construal level, you are thinking abstractly — about the 'why' of the thing. At a low construal, you are thinking in terms of concrete details — about the 'how' of the thing.

At a high construal, screwing in a light bulb is "bringing light to your daughter's room so she can read" or even "contributing to global warming." At a low construal, screwing in a light bulb is merely "rotating a sphere of glass." It is possible to think about our work in this way. At a low construal, you may describe your work in terms of tasks: A banker fills out forms, a programmer types code on a screen. At a high construal, you may talk about your work in terms of its broader purpose: a banker helps families get homes, a programmer invents tools to improve the lives of his customers. At a very high construal, work is described not only in terms of its

10

purpose, but in terms of how that purpose is personally significant to the worker (e.g., "I help families get homes, and I believe that everybody deserves a place they can call home, because I didn't have one when I was little."). In this example, the worker is perceiving their work as connected to helping others, to their own life history, and to their beliefs about what an ideal world looks like.

In their 2016 paper, Morrison et al. proposed this high-level construal of the work experience as a crucial element in both the evaluation of work as meaningful and the in-the-moment *experience* of work as meaningful. That is, if you find your work meaningful, by *any* definition or subjective interpretation of meaningfulness, you are construing your work at a high construal level because the process of relating experiences to your broader self and higher-order goals operates inherently at a high construal level.

If you do not find your work meaningful, you will construe it at a low construal level, because you just don't think of your work in terms of broader meaning. Construal Level Theory (CLT) allows us to approach meaningfulness from a different perspective. Instead of taking a philosophical, "here's what work means to humans" approach, Morrison et al.'s (2016) CLT perspective on meaningfulness talks about it in terms of the cognitive mechanics of how people interpret their work as meaningful. For this reason, it's potentially highly compatible and complementary to all extant theories of meaningful work. But, like most other definitions of meaningfulness, it is still just a theory. It needs to be tested empirically. As part of this study, we aim to do that.

Goal #1: Using Language to Inform Theory

The first goal of this study is to discover how people talk about their work when they find it meaningful (or meaningless) and draw implications from these language features about the

nature of meaningfulness. Natural language conveys a great deal of unfiltered information about how people think and feel (Mairesse, Walker, Mehl, & Moore, 2007). Recently, researchers have begun to capitalize on the psychological data latent in natural language by using machine learning techniques to discover linguistic indicators of various psychological states and traits (Kahn et al., 2016; Mairessee et al., 2007; Tausczik & Pennebaker, 2010). Researchers have found significant associations between language style and 'stable' individual differences like personality traits and basic values, and also with temporary psychological states like mood, emotion, and deception (Mairesse, Walker, Mehl, and Moore, 2007).

A particularly notable study by Rosenberg & Hirschberg (2005) found linguistic signals of leader charisma (charismatic people use more personal, first-person pronouns). They note at the start of their introduction that charisma is "more difficult to define than identify" as a justification for pursuing a lexical approach to measuring charisma — a line of thinking that that parallels our current struggle with meaningfulness. By studying charisma through the lens of language, Rosenberg and Hirschberg (2005) were able to discover that some degree of personal connection may be instrumental in creating an impression of charisma, which they note may not have been obvious to charisma researchers before their unique study. As this example shows, language analysis can help inform definitions of hard-to-define constructs. And as our present study will show, it proved to be a useful tool for understanding meaningfulness.

Linguistic signals of meaningfulness. How might a person talk about their work when they find it meaningful? In searching for previous research on linguistic signals of meaningfulness, I was only able to find *one* paper that explicitly proposed a relationship between certain language features and a sense of meaningfulness in work. Podolny et al. (2004) proposed — but did not test — a series of linguistic indicators of meaningfulness in work, using examples

12

from Terkel's (1974) popular book *Working* to illustrate their points.

    Personal Pronouns and Action Verbs. According to Podolny et al. (2004), meaningfulness appears not in what people say about their jobs, but in how they say it. Virtually the same statement could be spoken two different ways — one implying high meaningfulness, one implying low meaningfulness. When people find their work meaningful, Podolny et al. (2004) argue, they tend to talk about it using language that brings themselves closer to the work and to the people they work with. When people find their work meaningless, they tend to speak in a language that distances themselves from their work and their coworkers.

    As an example, consider the sentence "We're working hard to meet the deadline." versus "There's a lot of hard work going on here to meet the deadline." The first example contains the plural pronoun "we," which according to Podolny et al. (2004) indicates a close identification with both the author's work and their coworkers. The second sentence, by contrast, uses distancing language — "There's hard work going on here" — that doesn't even mention the author's self or others. The first sentence also contains verb-phrases (i.e., "working hard"), whereas the second example uses more "noun-like" language (e.g., "hard work"). Podolny et al. (2004) suggest that this 'speaking in nouns' language is also a form of self-distancing.

    At the basic level, Podolny et al. (2004) argued that highly meaningful work is discussed using first-person pronouns. Further, they suggest that first-person *plural* pronouns indicate that a person likely gets even more meaning from work because plural pronouns represent the connection of self to others. Finally, Podolny et al. (2004) suggested that a work story containing multiple, close-together sentences discussing work with first-person plural pronouns (e.g., gushing about "we") likely indicates that the worker gets very large amounts of meaning from their work.

As part of this study, I tested two of Podolny et al.'s (2004) proposed linguistic indicators of work meaningfulness: first-person pronouns and action verbs. Hypotheses 1-2 reflect these tests.

H1: People will describe their work using more first-person pronouns when they find their work meaningful.

H2: People will describe their work using more action verbs when they find their work meaningful.

Note that Podolny et al. (2004)'s notion of self-distance as a signal of low-meaningfulness has some tension with Morrison et al.'s (2016) proposition that workers will discuss work at a high-construal when it is meaningful, and at a low construal when it is meaningless. Construal Level Theory tightly links psychological distance with a construal level, where greater psychological distance always equals higher construal. It may appear, given this, that Podolny et al.'s (2004) notion of self-distance as a symptom of *low* meaning in work suggests that low meaningfulness will be associated with a *high* construal level.

This may prove to be confounding in practice, but theoretically, it is not contradictory. Podolny et al.'s (2004) notion of self-distance, loosely interpreted as "picturing your self as distant your meaningless job" is just one type of psychological distance that may inflate construal level. Looking at the *tasks* of your job from the perspective of your higher order goals also involves psychological distance between the concrete 'how' of the task and the broader 'why' of the task in which your meaningful state of mind is situated. In this way, Podolny et al.

14

(2004) proposed a psychological distance from the job itself, while Morrison et al. (2016) proposed a psychological distance from the tasks and concrete processes of that job.

Also, note that the act of "thinking about the self" as *either* near or far to the job is an act of high construal. This is evidenced in a brain imaging study by Van der Cruyssen et al. (2014) showing that the brain area associated with high construal states — the dorsal-medial pre-frontal cortex — is often referred to as the area responsible for thinking about the self.

It is entirely possible to arrive at a conclusion about whether you feel far or near to your work (as proposed by Podolny [2004]) *through* a high-construal metacognition. You can speak abstractly about your work being meaningful or meaningless, about yourself feeling tightly bound to it or detached from it, and all the while you are thinking about your work at a high level. The key differentiator, for my purposes here, will be whether the person talks about the *features* of their work using high-construal, abstract language --- not just *their relation* to their work.

Abstract words. Given the relationship proposed by Morrison et al. (2016) between high construal level and perceptions of work as meaningful, I expected high construal to be a necessary and usually-sufficient indicator that a participant finds their work meaningful. I propose that talking about work using low-construal, concrete language will usually be associated with low meaningfulness, because it suggests that the worker does not perceive a connection between his work and deeper values or purpose as salient (i.e., there is no "why" to the work). Conversely, talking about work features with abstract (high-construal) language should indicate a connection to "why" and should signal higher meaningfulness.

However, there are many ways one can operationalize 'high construal' in the context of language. For example, the word "fruit" is higher-construal (more abstract) than the word

"apple" (which is more concrete). However, the phrase "I saved the orphan with an apple" is higher-construal than the phrase "I gave the orphan a fruit" because the former gets at the 'why' of the action, while the latter phrase communicates only the concrete 'how' of the action (Vallacher & Wegner, 1989).

A more thorough rating of construal level language, designed to capture construal level at wider resolution (phrases, paragraphs, contexts) was conducted as part of this study and is described in detail below. However, I was also interested to see if construal level could be detected based simply on abstractness of individual words. To this end, Hypothesis 3 proposes that people will describe their work using more abstract individual words when they find their work more meaningful.

H3: People will describe their work using more abstract words when they find their work meaningful.

Positive Sentiment. There is a relative consensus in the theoretical literature on meaningful work that meaningfulness is positively-valenced (Lepisto & Pratt, 2016; Rosso et al., 2010). Generally, in a work context "meaningfulness" is discussed and thought of as a positive thing. Although it is possible for something to be full of dark meaning (e.g., visiting a concentration camp), people do not generally use the word that way, and it is unlikely that a worker will rate a job as high in meaningfulness that is meaningfully horrific to him/her. As such, I expect high meaning stories to be written in a positive tone and low meaning stories to convey a negative tone. There may be exceptions, as in the case of people who feel sadly "bound" to their highly meaningful work (see Bunderson and Thompson [2009]), but I expect these to be rare.

H4: People will describe their work using more positive sentiment when they find their work meaningful.

Summary. In summary, I expect that work stories rated as meaningful will include more frequent use of first-person (self-investing) pronouns, action verbs, abstract words, and positive sentiment.

Goal #2: Create a New, Natural Language-Based Measure of Work Meaningfulness

Simply discovering how people talk about their work when they find it meaningful would be useful for informing our understanding of what meaningfulness is. However, it was an important goal of this project to go beyond increasing our understanding of meaningfulness to also provide a tool to help other scholars take advantage of the additional insight provided by natural language to help measure meaningfulness in new ways. To this end, this study included as its second focus the creation of a new, language-based measure of work meaningfulness. This type of measure is called a Natural Language Processor.

A quick introduction to natural language processors. A Natural Language Processor is a computer algorithm that scans in a large amount of text and learns which features in that text (e.g., specific word usages, punctuation patterns, sentence length) are commonly associated with the presence of a targeted psychological construct. Once the language features for a construct are learned and built-in to the algorithm, the algorithm can then look for those features in new samples of language, and from that make an assessment about how much of the associated psychological construct is implied in the new text.

Perhaps the most 'famous' NLP measure (in terms of popular press coverage) is IBM's

17

Watson supercomputer, which can read-in a small amount of text (500 words, or about 6 emails) and produce statistically-significant estimates of the author's five-factor personality traits (with sub-facets) and a complete profile of the author's values — in less than one second (Mahmud, 2015; McCrae & Costa, 1987; Schwartz, 1994). And that's as of this writing. Watson is still learning. Within a year of Watson's launch, it learned to be more accurate and improved its efficiency by a factor of seven (meaning it needs less input text to arrive at the same predictive accuracy; Arnoux, 2016).

Watson is one example of a growing arsenal of such natural language measures being made available to researchers. As another prominent example, the popular Natural Language Processor called "LIWC" (Linguistic Inquiry and Word Count; pronounced "luke") can measure a battery of psychological constructs from language features, including need for achievement, time orientation, and analytical thinking (Pennebaker, Francis, & Booth, 2001).

An NLP measure of meaningfulness. As part of this study, I used machine learning techniques to create an NLP measure of meaningful work. The detailed construction of this algorithm is described below, but in essence: When run on a body of inputted text (specifically, a person describing their work), this algorithm will look for signals of work meaningfulness, and output a probability (between 0 and 1) that the author of the text finds their work meaningful. I expect these ratings to be significantly related to other, more traditional measures of meaningfulness.

H5: Estimates of meaningfulness produced by the natural language algorithm will correlate significantly with other measures of meaningfulness.

Construct validity check. I also expect that the meaningfulness scores outputted by the algorithm will relate significantly to phenomena that are known to co-occur with work meaningfulness. In particular, *affective commitment* is an established outcome of work meaningfulness (Jiang & Johnson, 2012). This relationship is both theoretically and empirically sound. Theoretically, you should want to stay in a job that you find highly meaningful. And empirical research has demonstrated that this commonsensical assertion seems to hold true, with commitment being one of the strongest outcomes of work meaningfulness (Jiang & Johnson, 2012; Seibert, Wang, and Courtright 2011).

H6: Ratings of meaningfulness produced by the natural language algorithm will correlate significantly with affective commitment.

Ultimately, this algorithm is designed to empower researchers to measure meaningfulness from samples of textual data (i.e., without having to ask people about meaningfulness via surveys and such). Hopefully, researchers will be able to use this new NLP measure to study meaningfulness within the many samples of natural language that occur organically in the modern workplace (e.g., cover pages, emails, open-ended survey responses, and interview transcripts).

Goal #3: Test the Role of Construal Level as a Potential Unifier

As discussed earlier in this paper, the theoretical space around meaningful work is scattered, siloed, and divergent. It is in dire need of unifying mechanisms. However, purely-theoretical unifying mechanisms may not be enough, as the space is already fat on theory and thin on empiricism. Indeed, reviews by both Bailey et al. (2016) and Rosso et al. (2010) lamented the lack of empirical work in the meaningfulness space and urged scholars to continue

forward to empirical testing, rather than creating more theory.

Thus, although Morrison et al.'s (2016) union of meaningfulness with Trope & Liberman's (2010) Construal Level Theory presents a promising candidate for unifying the meaningfulness space, until it is tested empirically it is effectively just 'yet another meaningfulness theory' in a literature full of untested unifying theories. This study aims to test Morrison et al.'s (2016) proposed relationship between construal level and meaningfulness and to (hopefully) establish it firmly as a *validated* candidate for unifying meaningful work theory. However, there are some unique challenges with testing construal level; challenges which this study aims to circumvent by assessing construal level through language rather than, say, through traditional Likert measurements.

The trouble with testing construal level. Although Construal Level Theory (CLT) represents potentially one of the most promising unifiers for meaningfulness theory, it cannot be measured easily with traditional Likert approaches.

Typically, when a new theory of work meaningfulness is introduced (e.g., Lips-Wiersma and Wright [2012]; Hackman and Oldham [1976]; Steger et al. [2012]), the researchers who introduce it create and test an accompanying Likert-scale measure for it, and then correlate responses on their Likert survey against some relevant outcome measures to demonstrate predictive and construct validity. These approaches work best to the extent that meaningfulness can be reduced to statements like "My work makes a positive impact on the lives of others" and "My work allows me to use many of my skills and talents" (as in Hackman and Oldham's [1976] Job Diagnostic Survey). A CLT perspective on meaningfulness would be incredibly difficult to test this way because construal level is difficult to operationalize and detect through traditional, multiple-choice scales.

20

We cannot ask a participant "Are you thinking about your work abstractly?" because the answer will always be "Well, *now* I am." This priming problem makes it difficult to design traditional, Likert-style scale prompts that get at construal level. The most popular measure of construal level is the scale created for Vallacher and Wegner (1987)'s Action Identification Theory (AIT), which is an extension of CLT. Vallacher and Wegner's (1987) measure is a forced-choice scale that describes various situations (e.g., "You're screwing in a light-bulb.") and then asks the participant to choose how they'd think of the action: with a concrete choice (e.g., "Rotating a sphere of glass.") and an abstract choice (e.g., "Bringing light to the darkness").

Vallacher and Wegner's (1987) Action Identification scale is possibly the best available approach to measuring construal level with traditional Likert scales, but there are still problems with it. First, there is an issue of demand characteristics. It's easy for the test-taker to detect an obvious relationship between the answer choices (i.e., that some are clearly abstract "higher-level" while others are specific and "low-level"). Second, the forced-choice format employed for Vallacher and Wegner's (1987) Action Identification scale limits the test taker to 2-4 predefined interpretations of a given action. Presumably, across all human minds in the world, there are more than 2 different ways to interpret screwing in a light bulb. And even more options than that when we're talking about something more complex than screwing in a light bulb, like a person's interpretation of their own job role. Here, it might be helpful to let people share their own interpretations of their work freeform, and then assess construal level post-hoc.

Testing construal level with language. This study tests the relationship between construal level and meaningfulness using natural language. That is, rather than asking participants to self-report their own construal level, participants provided a free-form, essay-like

description of their work, and a team of trained raters evaluated those essays and scored them for construal level. In line with Morrison et al.'s (2016) propositions, I expect to find a significant, positive relationship between these ratings of construal level and several measures of meaningfulness. In doing so, I hope to discover evidence that construal level could indeed be a common element in multiple theories of meaningfulness and to illustrate its power as a potential unifier.

H7: Construal level will have a significant, positive relationship with work meaningfulness.

Convergent validity check. If construal level is related to meaningfulness, then it should also be related to outcomes of meaningfulness. As discussed previously, affective commitment is a well-established outcome of meaningfulness, and so I expect that if high construal level signals high meaningfulness, then it will also relate to higher affective commitment.

H8: Construal level will have a significant, positive relationship with Affective Commitment.

METHODS

There were two overarching goals to this project: discover and measure how a sense of meaningful work shows up in people's language, and to test the relationship between meaningfulness and construal level. To achieve these goals, I needed three components: A

sample of language, a rating of meaningfulness, and a rating of construal level — from people who work.

In order to build the desired computer algorithm to predict meaningfulness from language (called a 'text classifier' because it classifies text as either meaningful or not meaningful), I needed to collect a sufficiently large sample of work stories to be able to 'train' the algorithm on some texts, and then test it on a different sample of texts that the algorithm had never seen before.

Participants

Participants were **N = 194** Mechanical Turk users (n = 76 male, n = 118 female) who identified as being employed full-time. Participants were asked to complete a survey about their work. Generally, for text classification algorithms like the one developed in this study, a sample size of between 80 and 560 texts is recommended for satisfactory text classifier performance (Figueroa, Zen-Treitler, Kandula, & Ngo, 2012). Given budget constraints, I aimed for the mid-lower-end of this recommendation (194 stories).

Compensation. Given that this survey asked about work attitudes (i.e., meaningfulness), there was some concern that the pay rate itself could shift attitudes. I aimed for a pay rate that fairly and precisely matched the task requirements: for one, to use the money efficiently, but also so as not to add contaminating positive/negative affect due to feelings of being under/over-benefited. Pilot testing was used to arrive at an ideal pay of $7.50 for completing this study. Lower payments than this resulted in comments suggesting the pay was too low, with one participant commenting (at $6.00) that "The pay rate was okay for the time, with all that writing though, I honestly think $7 is a better price point." At the final pay rate of $7.50, comments like "pretty fair" or "pay rate was good" were common.

Collected Data

      Work story. Each participant was prompted to write 500 words (about a page) about their work. These work stories are the key piece of the data; the primary focus of analysis. A goal of this project was to collect work stories that were reasonably naturalistic and generic; that is, to capture people speaking as they normally would about their job as if discussing their work at a party, or in an interview, or in response to an open-ended question on an employee survey. This is as opposed to speaking about meaningfulness directly. It was important for the generalizability of the measure that the participants not be too primed to speak about any specific aspect of their work.

      Story prompt. Recall that this project proposes that fluctuations in construal level may indicate meaningfulness. Priming a particular construal level would confound the endeavor of testing for natural fluctuations in construal level. Thus, the prompt for the story collection was designed to be natural and neutral enough to prime people to speak broadly enough about their work that they were tacitly encouraged to make overall judgments like meaningfulness, without directly asking them to talk about the meaningfulness of their work or otherwise artificially inflating (or deflating) their construal level.

      For example, a prompt like "Why do you do your work?" would have likely primed a high construal level, as asking 'why' has been used in other studies to prime high construal (Schwartz, Eyal, & Tamir, 2018). Conversely, a prompt like "Describe your job role and the tasks you need to complete" would have likely primed a low construal level, as attending to low-level details is a feature of low-construal (Trope & Liberman, 2010).

      Several different prompts were piloted to see which seemed to elicit the best balance of abstraction in the story texts. Surprisingly, many initial prompts resulted in insufficient variance

in meaningfulness ratings. That is, *most people found their work meaningful*. In one pilot, everyone found their work highly meaningful.

This difficulty finding meaning*less*ness was surprising and seems to suggest that most people, upon reflection, find their work at least somewhat meaningful. This is in itself interesting and perhaps deserving of further study, as it is consistent with Frankl's (1962) proposition that humans have a "will to meaning." Or perhaps, consistent with the construal level relationship proposed here, simply asking people to reflect on their work at all — no matter how the question is phrased — prompts people to think of their work at a distance, abstractly, thus raising their construal level and makes it seem more meaningful. Whatever the explanation, this lack of variability in meaningfulness scores in the pilots presented a potential challenge for trying to analyze the data. It would be difficult to design an algorithm to separate high and low meaningfulness stories if they were all too some extent high in meaningfulness.

Ultimately, after several pilots and consulting with several advisors, the prompt "In 500 words, tell me about your work," combined with a more severely-worded prompts for the primary meaningfulness measures ("Think of the most meaningful work you can imagine for yourself. Now, how meaningful do you find your current work?" and "Overall, I find my work *very* meaningful."), seemed to be the most neutral and capable of eliciting acceptable variability in meaningfulness scores.

These final prompts were chosen by 'eyeballing' the variance in meaningfulness scores in each pilot and judging them as (at last) satisfactory (seeing noticeably more low-meaning scores in the final pilot). I realized after data collection that a better approach would have been to choose the best prompt by formally comparing the mean meaningfulness scores of each. I did perform this analysis retrospectively and found that (luckily) the final prompt used in the full

data collection exhibited the *second*-best variability in meaningfulness scores in the pilots. The first-best prompt was also the smallest pilot, so it's not clear that it was really the best prompt or simply chance that lead to its slightly superior variability in meaningfulness.

Despite apparent progress in pilot testing (you can see the means approaching the actual scale midpoint of 3 as the prompts were iterated with each pilot), the means in the final study were still weighted heavily towards overall high-meaning. It's possible that none of this prompt experimentation made a difference and I just caught some extra low-meaning people coincidentally in some pilots. Table 1 below lists all prompts and means. Note that the continuous meaningfulness ratings were on a 1-7 scale, and the binary meaningfulness ratings were on a 0-1 scale.

Table 1. Pilot work story prompts.

	Story prompt	Continuous scale criteria text	Binary criteria text	Mean Continuous Meaningfulness (Binary Meaningfulness)
Pilot 1	In 500 words, tell me about your work.	"I find my work meaningful."	n/a	5.5 (na)
Pilot 2	In 500 words, tell me about your work.	"I find my work personally meaningful"	n/a	5.0 (na)
Pilot 3	Imagine you are at a party, and somebody asks you "What do you do for a living?" How would you respond?	I find my work meaningful.	Overall, do you find your work very meaningful?	5.7 (.86)
Pilot 4	In 500 words, tell me about your work. (Sunday)	I find my work meaningful.	Overall, do you find your work very meaningful?	6.0 (1.00)
Pilot 5	In 500 words, tell me all about your work.	I find my work very meaningful.	Overall, do you find your work very meaningful?	5.6 (.82)
Pilot 6	In 500 words, tell me all about your work.	How meaningful do you find your work? [Meaningless, Not very meaningful, Generally meaningful to The most meaningful work I've ever done]	Overall, do you find your work very meaningful?	5.6 (1.00)

Table 1 (cont'd)

Pilot 7	In 500 words, tell me all about your work.	How meaningful do you find your work? (vs Most meaningful work I can imagine)	Overall, do you find your work very meaningful?	4.8 (.81)
Pilot 8	In 500 words, tell me all about your work.	Added a definition of work meaningfulness to consider.	Overall, do you find your work very meaningful?	3.2 (.55)
Pilot 9	In 500 words, tell me all about your work.	Added "Think of the most meaningful work you can imagine for yourself. Now, how meaningful do you find your current work?"	Overall, do you find your work very meaningful?	4.55 (.77)
Final	(unchanged)	(unchanged)	(unchanged)	5.1 (.85)

Explicit meaningfulness story. Concerned that a neutral story prompt might fail to pick up *any* meaningfulness in tone or language, I collected a second mini-essay from each participant asking them to explicitly talk about the meaningfulness of their work. Participants were asked to write 250 words in response to the prompt "Overall, do you find your work very meaningful? Why or why not?" Although I ultimately found sufficient patterns in the generic story, these explicit meaningfulness stories proved highly useful for coming up with potential text themes to look for.

Single-item meaningfulness catchall. I included a single, straightforward Likert-scale item to capture meaningfulness broadly on a continuous scale. Although the original intent was to provide a theory-agnostic catchall for capturing meaningfulness broadly, during the process of combatting the lack of variability in meaningfulness ratings (discussed above), I ultimately incorporated a definition of meaningfulness in this prompt. The full item starts with a definition of meaningfulness, drawn from Morrison, Walker, and DeShon (2016): "The idea of 'work meaningfulness' means different things to different people. For our purposes here, we say that work is meaningful when it feels connected to your deepest values, goals, and needs." Note that the word phrase 'deepest values' was used in place of the original definition's "higher-order values" for interpretability.

Before completing the scale, participants were asked the question "Think of the most meaningful work you can imagine for yourself. Now, how meaningful do you find your current work?" Responses to this item were on a 7-point Likert scale, with the following scale points:

- 0 - Meaningless
- 1 - Not very meaningful
- 2

- 3 - Mostly meaningful

- 4

- 5

- 6 - The most meaningful work I can imagine for myself.

Self-report measures of meaningfulness. After providing the work story and explicit meaningfulness story, participants completed several traditional, Likert measures of work meaningfulness. The scores from these traditional measures were used as criteria together with human ratings (other-ratings, described below) of meaningfulness to inform the predictive algorithm. That is, the algorithm was developed to discover indicators in natural language that were associated with high and low scores on many different measures of work meaningfulness. My goal was for this NLP measure to not simply predict as well as any *one* existing meaningfulness measure, otherwise it would essentially be an NLP version of that measure, but instead to predict as well as *multiple* popular assessments of meaningfulness.

The Work and Meaning Inventory. Participants completed the current most-cited Likert measure for assessing work meaningfulness: The Work and Meaning Inventory (Steger, Dik, & Duffy, 2012). This is a 10-item measure of experienced meaningfulness in work, divided into three factors: Greater-good motivations ("the degree to which people see that their effort at work makes a positive contribution and benefits others or society"), Meaning-making (i.e., seeing your work as something that helps you make sense of your life), and Positive meaning ("The degree to which people find their work to hold personal meaning, significance, or purpose").

The Comprehensive Meaningful Work Scale. Several prominent reviews of meaningfulness (e.g, Rosso et al., 2010; Lips-Wiersma and Wright, 2012; Barrick et al., 2012) have proposed that there may be many different pathways to finding a sense of meaning in work,

30

and that there may even be individual differences in which pathways actually provide meaning for different people (e.g., a mathematician may find meaning in self-efficacy, a nun may find meaning in self-transcendence).

Although Rosso et al.'s (2010) pathway model seems to be the most dominant, based on citation counts, at present there is no measure through which I could capture the pathways to meaningfulness they propose. Lips-Wiersma and Wright's (2012) Comprehensive Meaningful Work Scale, however, is based on a pathway model that is so similar to Rosso et al.'s (2010) model that Rosso et al. recently issued a corrigendum apologizing for not acknowledging it more seriously in their review (Rosso, Dekas, and Wrezniewski, 2011).

Including Lips-Wiersma and Wright's (2012)'s Comprehensive Meaningful Work Scale allowed me to collect information on *where* (in terms of existing theoretical categories) participants found meaning in their work. Again, this helped inform the themes I looked for in the text to assess overall meaningfulness and could allow future versions of the measure introduced here to use this criterion to assess both overall meaningfulness and which pathway(s) a particular person finds meaningfulness through.

Affective commitment. In addition to the battery of meaningfulness measures, participants also completed a short measure of affective commitment (Allen and Meyer, 1990). This served as a construct validity check, as affective commitment has been shown to correlate with meaningfulness (Jiang & Johnson, 2012).

Binary meaningfulness question. Suspecting that my relatively small sample size would limit the scale points my algorithm could predict (discussed in detail later), I included a simple, two-class binary meaningfulness question. Participants were asked the binary question "Overall, is your work very meaningful? Yes/No". This binary "yes/no" meaningfulness question served

31

as a validation check on the human ratings of meaningfulness (discussed below), and as a backup target criterion for the algorithm.

Human Ratings

One of the concerns we discussed at the onset of this project was that asking for self-reports of meaningfulness could be tricky because each participant might define 'meaningfulness' differently for themselves. For example, anecdotally it seems like many lay-people equate the notion of meaningful work with *altruistic* work, even though in the literature altruism is only one of many pathways to meaningfulness (Rosso et al., 2010). To address these concerns about definitional consistency, I trained a team of raters to rate each story for meaningfulness according to a consistent definition.

I also trained these raters to assess construal level as part of this study's attempt to test the relationship between construal level and meaningfulness. In sum, raters produced judgments of both meaningfulness and construal level for each story. The other-rated scores of meaningfulness were tested in concert with the self-report scores to serve as potential prediction targets for the natural language algorithm.

Other-rated meaningfulness. Each rater read the Morrison, Walker, and DeShon (2016) paper which proposes an integrative definition of meaningfulness. Following this, they were asked to write two 500 word essays about work (similar to those collected from participants): one essay about the most meaningful work they'd ever done, and one about the most meaningless work they'd ever done. They were also asked to include a paragraph after each story introspecting about how they talked about each work experience. Finally, raters were walked through examples of high and low meaningfulness stories (from pilots).

Raters rated each work story for meaningfulness on a 1 to 7-point scale, and also made an

overall (binary) classification decision for each story as either meaningful or not meaningful. Raters were also asked to provide notes explaining their ratings.

Other-rated construal level. In addition to rating for general meaningfulness, raters also rated stories for their level of construal. That is, the extent to which the participants talked about their work abstractly or concretely (i.e., spoke about the forest or got lost in the trees). Collecting a separate rating of construal level was not directly related to my goal of creating a Natural Language Measure for meaningfulness (I could have created the meaningfulness measure without it). However, I did this opportunistically, as this project provided an opportunity to test the theoretical relationship between construal level and meaningfulness. I expected that level of construal would be correlated with the amount of meaningfulness a person indicated getting from their work, and with a numeric rating of construal level provided by raters I could compare it to ratings of meaningfulness and test the relationship.

How construal level was rated. Raters were educated on the notion of construal level with reading assignments and response papers. Each rater read the original Trope and Liberman (2010) paper on construal level and wrote a response paper on the concept. They were then walked through examples of spotting high/low construal language and themes in the work stories (drawn from pilots). Raters provided two ratings of construal level for each work story: a 1-7 rating (from concrete to abstract), and a binary overall classification of either high or low construal.

Machine Ratings

Several of the hypotheses required the calculation of endogenous statistics based on features of the collected story texts, to be used in hypothesis testing.

Abstract words. To test Hypothesis 3 ("People will describe their work using more

abstract words when they find their work meaningful."), I used a publicly-available dictionary by Brysbaert, Warriner, and Kuperman (2014) that includes 40,000 English words rated for concreteness of language. Using this dictionary, I was able to 'look up' the concreteness score of each word used in a particular work story (provided that the word was included in the dictionary), then add those scores together to form a cumulative concreteness score for the entire story. I then reversed-scored this concreteness score, so that I could treat it as an abstraction score. In summary, the abstract score for a given story was calculated by summing the abstractness (reverse-scored concreteness) scores for each word in the story and then dividing that sum by the total number of words in that story.

Sentiment. Hypothesis 4 ("People will describe their work using more positive sentiment when they find their work meaningful.") involved testing for positive sentiment. To accomplish this, I computed sentiment scores for each work story using the SentimentIntensityAnalyzer function included within Python's Natural Language Toolkit (NLTK) package.

Side note: This SentimentIntensityAnalyzer software tool was built using a similar process I used here to construct my meaningfulness NLP measure: a massive sample of text was collected and scored by human raters for positive/negative sentiment. These text and sentiment score examples were then used to train a machine learning model, which could then be applied to new bodies of text (like these work stories) to evaluate their sentiment.

Sentiment scores were computed for each work story, and these sentiment scores were tested for their relationship with all collected measures of meaningfulness using a Pearson correlation test. The SentimentIntensityAnalyzer package returns four scores for sentiment: an amount of positive sentiment, amount of negative sentiment, amount of neutral sentiment, and compound sentiment intensity. Thus, I was able to test the correlation between meaning and

positive sentiment directly (versus other packages, which would require testing against 'overall sentiment' on a continuum from negative to neutral to positive).

Parts of speech. To test Podolny et al.'s (2004) assertions that highly-meaningful stories will involve more first-person possessive pronouns and action verbs (Hypotheses 1-2), I used Python's NLTK package to 'tag' every word in every work story with its part-of-speech, using the standard Penn Treebank part-of-speech tagging convention (Santorini, 1990). For example, the word 'work' became either 'work_VERB' or 'work_NOUN' depending on how it was used. The part of speech interpreter included in Python's NLTK has some ability to be context-aware like this and attempts to determine a word's part of speech based on the word itself *and* where/how it appears in the text. Note that the actual part-of-speech tags are more precise than simply "VERB," and use special abbreviations to denote particular parts of speech (e.g., VBP = Verb, non-3rd person singular present).

I counted the occurrence of each part of speech proposed in Podolny et al.'s (2004) in each story. In terms of the Penn Treebank's part-of-speech abbreviations, for action verbs I counted occurrences of VBPs (Verb, non-3rd person singular present) and for first-person possessive pronouns I counted PRP and PRP$s (personal pronouns and possessive pronouns). Thus, a count score (weighted by story word count) was computed for each of the targeted parts of speech for each story. Part-of-speech scores were tested for their relationship with all collected measures of meaningfulness using a Pearson correlation test.

Developing the Algorithm

Again, the second goal of this study was to develop an algorithm that looks for a set of language features that — when found in a description of work — reliably predict whether a person finds their work meaningful or not. NLP measure development involves a great deal of

discovery and trial and error, but I will explain here the general process that I went through to discover these features.

Choosing the optimization parameter. In machine learning, an optimization parameter gives the learning algorithm a goal to shoot for, or more accurately a 'success' criteria. The simplest and most common approach to optimization parameters is to use a binary optimization parameter: a zero is failure, a one is success. I collected two of these binary parameters: A binary self-report "Overall, do you find your work very meaningful?", and a binary other-report created by raters (i.e., "Does this person find their work meaningful? Yes/No").

At the study onset, there was interest in using the human ratings of meaningfulness as my optimization parameter. However, I discovered that these human ratings had poor correlations with established self-report measures of meaningfulness. Thus, doubting their construct validity, I chose to use the binary self-report of meaningfulness instead, as this had satisfactory correlations with all self-reported measures of meaningfulness as well as the construct validity check measure (affective commitment). See Table 2 for a summary of the considered optimization parameters.

Table 2. Correlations between potential optimization parameters and collected measures of meaningfulness and commitment. All correlations are significant with p < .001

Optimization Parameter	Self-Report Continuous Meaningfulness[1]	WAMI[2]	Affective Commitment
Human-rated binary meaningfulness	.35	.35	.32
Self-report binary meaningfulness	.69	.78	.57

[1.] **Self-reported Continuous Meaningfulness** = "How meaningful do you find your current work?" From "Meaningless" to "The most meaningful work I can imagine for myself

[2.] **WAMI** = Work and Meaning Inventory (Steger et al., 2012)

Not enough power for a continuous optimization parameter. Although I collected continuous measures of meaningfulness (on a 1-7 scale), training a text classification algorithm to predict each point on a 7 point scale would have required (according to best practices) at least 40-280 participants *per scale point* to achieve good accuracy, so a 7-point scale would have necessitated a sample size of n = 280-1960 — and that assumes that the variance would be evenly distributed, which as we've seen in this study seems tricky to achieve when measuring meaningfulness (Figueroa et al., 2012). Even still, I did attempt to point the algorithm at the collected continuous parameters, just to see what would happen. As expected, it's accuracy rate plummeted abysmally (though interestingly, still better than chance).

Creating a training set. Following best practices in NLP measure development, I randomly assigned half of my work stories to a "training set" and the other half to a "testing set." I used the "training set" to develop the measure. Once a set of language features was discovered that predicted meaningfulness scores well on the training set, I tested these features on the "testing set." Developing the measure on a separate set of stories, and then testing the measure on a "fresh" set of stories in this way helps ensure the measure's generalizability and guard against over-fitting the features to the data (Shan, Wang, Chen, 2015).

Version 1: Bag of words. In any machine learning model, one must create a feature vector. The purpose of a feature vector is to quantify features of your sample that may originally be nominal or descriptive. For example, a simple feature vector may contain counts of occurrences of the words "happy" and "sad" contained in a text (this approach of searching simply for word occurrences is called a "bag of words" approach). Initially, I used Python's SciKit Learn package (a well-regarded set of machine learning functions for Python) to treat all words in all work stories as features (with the key being the word, and the value being the

38

number of occurrences of that word in the document). I also added phrases of 2 and 3-word length to this feature vector. I then used an XGBoost classifier to determine the most predictive words and phrases in the text.

Prediction scores from this model were good. Generally, this model could predict self-reported meaningfulness with an accuracy of 82%. However, the discovered 'most predictive' features often did not seem to be theoretically meaningful. Although this model was statistically sensible, it did not seem to shed much theoretical light on the meaningfulness construct as I'd hoped. Additionally, by this point, I had anecdotally observed while reading the collected work stories several seemingly clear differentiating features between high-meaning and low-meaning stories that the simple bag-of-words model wasn't able to incorporate easily (because they were more complex patterns, like 'first words contain...'). Desiring a more defensible, more flexible, more theoretically-driven predictive model than the 'black box' bag of words I had created, I switched approaches.

Version 2: A theory-driven model. My first algorithm employed generic machine learning tools to build a text classifier. In my second and ultimately more successful algorithm, I used tools designed specifically for text classification problems (versus machine learning in general). Rather than treating all words in all stories as potential predictive features, this new algorithm was built using only a handful of highly-discriminating features (i.e., words and phrases that were conspicuously present in meaningful stories, and conspicuously absent from meaningless stories). A NaiveBayes machine learning classifier then tested the features I suggested to it to determine how powerful they were in predicting ratings of meaningfulness.

Creating a search function Abandoning automatic feature detection forced me to use a more manual search process for potential language features that could signal high or low-

meaningfulness. To aid in this search, I created a Python function to accept a regular expression (regex) as a pattern in addition to simple words and phrases. Regular expressions are a somewhat-universal programming syntax for defining advanced search patterns, and incorporating them allowed me to investigate more complex patterns as potential features (e.g., phrases in a certain position, and phrases with some wording variability; Thompson, 1968).

This regex investigator function searched for whatever regular expression pattern I passed it in each story, and returned the number of occurrences of that pattern in high-meaning stories, relative to the number of occurrences in low-meaning stories. This ratio represented a difference delta of sorts, either positive or negative. Individual story length was also included in this calculation, to ensure that the search didn't conclude that a pattern was more frequent, simply because a story was longer.

This function was in some ways a manual version of what typical bag-of-words feature searches do automatically, but made more flexible by regular expressions, and more preliminary in its conclusions.

Searching for features. Armed with the ability to get a quick 'discrimination ratio' for any given pattern, I began reading the stories and searching for patterns inspired by meaningful work theory. When I discovered a highly-discriminating pattern, I added it to a custom, from-scratch vector of features, and retested the classification algorithm for accuracy. Features that improved the predictive accuracy score were retained; features that did not improve the accuracy score were excluded. Note: the specific 'most predictive' new language features discovered through this process are discussed in detail in the Results section.

Naive Bayes classifier. While my first algorithm used XGBoost, a well-regarded 'boosted trees' classifier widely recommended for overall machine learning tasks, this second

algorithm used a Naive Bayes (NB) classifier, which seems to be used more commonly in text classification projects in particular (Naive Bayes Classifier, 2018). Additionally, because the NB classifier is 'baked-in' to Python's NLTK package, it incorporates several features specifically designed to aid the development of text classification algorithms that were absent from more general Machine Learning packages like XGboost.

Cross-validation. Once I had exhausted ideas for new, theoretically-inspired patterns to look for in the text, and was also satisfied with the accuracy (predictive validity) of the algorithm, I performed a final validation of the algorithm to ensure that its accuracy was relatively stable. As mentioned previously, I split my story data (according to standard practice) into test and training sets, which comprised roughly 66% and 34% of the data, respectively. The catch here is that the split is (intentionally) random. Each time the algorithm is run, different stories are used for testing and training. For this reason, classification algorithms vary in accuracy depending on what training and test data they use.

To account for this variability issue and get a 'stable' estimate of how accurately an algorithm will be able to predict when pointed at new data, standard practice in machine learning is to validate classification algorithms using a "K-Fold" test (K-Fold Cross-Validation, 2018). In a K-Fold test, the data is split into several sub-samples, then re-trained and re-tested on each sample (Brownlee, 2018). The average accuracy of each fold is then reported as the overall 'skill' of the classifier.

Construal Level and Meaningfulness

To test the relationship between other-rated construal level and meaningfulness, I performed a Pearson correlation test between the averaged other-rated Construal Level score for

41

each work story and all collected measures of meaningfulness. Recall that testing construal against multiple measures was not an act of just "throwing it at the wall and seeing what sticks." This multi-criteria test was central to my argument that construal level is a universal component of meaningfulness: that is, it *should* correlate significantly with *all* validated measures of meaningfulness.

RESULTS

Table 3 displays the correlations between all study variables. In short, the language-related hypotheses were partially supported, the algorithm-related hypotheses were fully supported, and the construal level hypotheses were fully supported. Or, only about half of the language features I expected (in advance) to be signals of meaningfulness worked, but I discovered some brand-new language features that predict meaningfulness, and construal level and meaningfulness seem to be related.

Table 3. Correlations of all variables.

		1	2	3	4	5	6	7	8	9	10	11	12
1	Other-rated meaningfulness	--											
2	WAMI[1]	.40**	--										
3	CWMS[2]	.34**	.81**	--									
4	Self-reported meaningfulness	.39**	.79**	.66**	--								
5	Self-reported meaningfulness (binary)	.26**	.78**	.62**	.70**	--							
6	CL	.74**	.21**	.16**	.20*	.08	--						
7	Personal Pronouns	.31**	-.04	-.02	-.03	-.05	.48**	--					
8	Action verbs	.24**	-.02	.01	-.02	-.06	.26**	.63**	--				
9	Abstract Language	.05	.11	.09	.02	.13	.09	-.01	-.02	--			
10	Positive Sentiment	.52**	.20*	.17*	.22**	.11	.57**	.34**	.23**	.16*	--		
11	Affective Commitment	.36**	.74**	.79**	.62**	.57**	.21**	-.00	-.02	-.01	.21**	--	
12	"I am a…"	.12	.19**	.20*	.31**	.18*	.18*	-.09	.02	.03	-.02	.23	--
13	Algorithm-predicted meaning.	.10	.29**	.26**	.35**	.31**	.09	-.10	-.17*	-.03	.01	.24**	.24**

* p < .05, ** p < .01, [1] **WAMI** = Work And Meaning Inventory (Steger et al., 2012), [2]**CWMS** = Comprehensive Work as Meaning Inventory (Lips-Wiersma & Wright, 2012).

Goal 1: Discovering the Language of Meaningful Work

      First-person pronouns and action verbs. Recall that Podolny et al. (2004) interpreted the *apparent* meaningfulness conveyed in work stories collected in Studs Terkel's (1972) book *Working*. That is, they made a judgment about how much meaning each story author seemed to get from their work, and then those judgments were used to inform their proposals about the language features associated with highly-meaningful (or meaningless) work stories. They could not ask the authors of the stories how meaningful they found their work, so they attempted to judge it themselves. Thus effectively, Podolny et al.'s (2004) propositions could be taken as saying "these are the language features that make a person *sound like* they get a lot of meaning from their work."

      In this study, Hypothesis 1 and Hypothesis 2 tested two of the language features that Podolny et al.'s (2004) suggested would be related to a sense of meaningfulness in work: personal pronoun use and action verb use.

      H1: People will describe their work using more first-person pronouns when they find their work meaningful.

      H2: People will describe their work using more action verbs when they find their work meaningful.

      I tested each of these hypotheses by performing a Pearson's correlation test on the relationship between personal pronouns and (separately) action verbs with all collected measures of meaningfulness. Tellingly, the strongest (and only significant) correlations were found

45

between Podolny et al.'s (2004) suggested language features and *other*-ratings of meaningfulness. Partially supporting H1 and H2, personal pronouns and action verbs were both significantly related to other-ratings of meaningfulness (personal pronouns: r = .31, p < .001; action verbs: r = .24, p < .001). None of the correlations between personal-pronouns/action verbs and any *self-report* of meaningfulness were significant.

These findings seem to support Podolny et al.'s (2004) notion that first-person pronouns and action verbs are related to a work story *sounding* meaningful to external human raters, but fail to support Podolny et al.'s (2004) suggestion that these features relate to 'actual', self-reported feelings of meaningfulness.

Abstract language. Consistent with the relationship proposed in this study between construal-level and meaningfulness, Hypothesis 3 proposed that people would use more abstract words (at an individual word-level resolution) when discussing work that they found meaningful.

H3: People will describe their work using more abstract words when they find their work meaningful.

Using the concreteness dictionary provided by Brysbaert et al. (2014) to compute an abstract word score for each work story, I performed a Pearson correlation test on these abstraction scores to determine whether they related significantly to any of the collected other/self-report measures of meaningfulness. No relationships were significant with any measures of meaningfulness. Therefore, Hypothesis 3 was not supported. People do not appear to use more abstract individual word choices when they find their work meaningful. Given the support found for the related construal-level hypotheses (described below), it's possible that this

46

finding reflected the appropriateness (or lack thereof) of the concreteness dictionary approach employed for this step more than an underlying theoretical failure.

Positive sentiment. Hypothesis 4 tested the relationship between sentiment and meaningfulness.

H4: People will describe their work using more positive sentiment when they find their work meaningful.

To test this relationship, I used a sentiment analysis tool included in Python's Natural Language Toolkit, computed a sentiment score for each story, and tested those scores against all collected measures of meaningfulness using a Pearson correlation test. A significant relationship was found between positive sentiment and all collected measures of meaningfulness (see Table 4). Therefore, Hypothesis 4 was fully supported. However, it is notable that positive sentiment showed only small correlations with 'actual' self-reported meaningfulness, suggesting that positive sentiment may not be necessary for meaningfulness. However, positive sentiment seems to have a large relationship (**.52**) with making work stories sound meaningful to external human raters.

Table 4. The relationship between positive sentiment and meaningfulness.

	WAMI[1]	Single-Item Meaningfulness Catchall	CWMS[2]	Other-rated Meaningfulness
Positive Sentiment	.20*	.22**	.17*	.52***

*$p < .05$ **$p < .01$ ***$p < .001$, [1]**WAMI** = Work And Meaning Inventory (Steger et al., 2012),

[2]**CWMS** = Comprehensive Work as Meaning Inventory (Lips-Wiersma & Wright, 2012).

All-new linguistic signals of meaningfulness. Recall that a primary goal of this study was not just to test existing ideas about how meaningfulness appears in language (i.e., those from Podolony et al.; 2004), but to discover brand-new signals of meaningfulness in language with the aid of machine learning tools. As mentioned above, the process of discovering new linguistic signals of meaningfulness was largely one of trial-and-error. I tried adding over 100 different language patterns to the algorithm before arriving at a relatively simple predictive model with good accuracy.

What follows is a description of the language features that worked. That is, language features that through trial-and-error were found to have the largest positive effect on the algorithm's ability to predict meaningfulness ratings.

A note on communicating the significance of these features: In the case of one extremely strong and reliable feature, I was able to illustrate its relationship as a significant correlation. However, for the remaining new language features, although they improved the accuracy of the algorithm, they did not occur commonly enough overall to show even a weak correlation with meaningfulness measures on their own. Trying to illustrate the validity of these language features using correlation tests would be like trying to correlate the color "green" with a scale that ranged from "Orange" to "Apple". If the fruit is green, it's definitely an apple rather than an orange, so "color = green" is a good predictor of apple-ness, but enough apples are red to muffle the correlation between "green" and "apple" in most apple samples.

Thus, I have indicated the predictive power of each of the new language features discovered by listing their likelihood ratios as determined by the ML algorithm (explained in detail below). All that said, here are the new linguistic signals of meaningfulness uncovered by this study, and theoretical support to help explain why they might help predict meaningfulness as

49

well as they do. Note that the predictive power ratios for these features (displayed as Machine Learning 'likelihood ratios') are listed in Table 6 and Table 7.

Identity statements. When introducing each of their work stories, participants used a variety of "I-statements". For example, "I work at...", "I work as...", etc. There was a manageably-finite number of these intros (e.g., there were 34 different first two-word combinations, out of 194 stories). One of the most obvious and face-valid language differences between low-meaning and high-meaning stories was that stories with lower meaningfulness scores (everything less than the highest meaningfulness allowed by the continuous single-item scale) tended to begin their stories with variations like "I work at…, I work for..., I work as…". However, stories with extremely high self-reported meaningfulness (max scale value of 7, labeled "the most meaningful work I can imagine for myself") were much more likely to begin with the words "I am...".

To help you understand the starkness of this difference, only 12% of stories with a self-reported meaningfulness of anything less than 7 (the highest possible) began with "I am a...". In comparison, 42% of the stories with the highest-possible meaningfulness score (7) began with "I am…".

This could suggest that those who find their work extremely meaningful incorporate their work more deeply into their identity. Interestingly, the notion of identity investment being associated with meaningful work is central to Podolny's et al. (2004) propositions. This seems to suggest that Podolny et al.'s (2004) overall theory was highly prescient, but perhaps (based on the findings in this study) the particular language features they suggested to operationalize their theory may need to be refined further.

The specific pattern used to teach the algorithm how to look for these types of 'identity

statements' at the beginning of the work story is below. As this pattern was consistent enough to show up in a correlation test, the correlation between the appearance of this pattern and self/other-reports of meaningfulness, as well as with affective commitment (the construct validity check) is shown in Table 5.

To illustrate how especially powerful this "I am a…" language feature is for predicting *extremes* of meaning, I have also included (in Table 5) correlations between the identity statements pattern and meaningfulness in a polarized version of the Work Stories Corpus that included only extremely high ("The most meaningful work I can imagine for myself") and extremely low ("Meaningless" and "Not very meaningful") self-reported meaningfulness stories. In this polarized dataset, the correlation between "I am" language and meaningfulness *doubles* for most measures of meaningfulness.

Regex search pattern for 'identity words':

```
^(I am a |I'm a |I am an |I'm an )
```

Table 5. The relationship between "I am" language and meaningfulness.

	WAMI[1]	Single-Item Meaningfulness Catchall	CWMS[2]	Other-rated Meaningfulness
Correlation with "I am a" language. (full dataset)	.19**	.31***	.20**	.12 (ns)
Correlation with "I am a" language. (polarized dataset)	.41***	.43***	.40***	.36**

p < .01 *** p < .001, [1]WAMI** = Work And Meaning Inventory (Steger et al., 2012), [2]**CWMS**

= Comprehensive Work as Meaning Inventory (Lips-Wiersma & Wright, 2012).

Temporariness Words. If you ask somebody what they do for a living, and they reply "Well currently..." it's probably safe to assume that they don't plan on staying in their current job, perhaps because they don't find it satisfying or meaningful. I noticed that some of the work stories started with statements like this, and so I created a search pattern to look for signals of temporariness, or intention not to remain in the job. This feature pattern helped the algorithm identify low meaningfulness stories.

Regex search pattern for 'temporariness words':

```
^((I am currently)\s[^(a)])|^(My current job)|I currently work\s[^(as)]
```

Work centrality language. Work centrality is a concept often discussed as being related to meaningfulness (Rosso et al., 2010). Workers with high work centrality consider their work a centrally-important aspect of their life. Positing that workers with higher meaning might have higher work centrality and thus higher preoccupation with their work, I created a pattern to look for potential indicators of work preoccupation. This pattern is small, and likely does not cover the full range of words likely associated with work centrality (this could be developed further in a future project). Nonetheless, it did improve the accuracy of the algorithm consistently, but only by a small amount (see Table 6 and Table 7).

Regex search pattern for 'work centrality words':

```
worry|care
```

Pace. Noticing that some of the low-meaning stories seemed to lament the pace of their work, I decided to incorporate the word "pace" into the model. It turned out to be a good

differentiator, showing roughly a 2:1 ratio of occurrences in low-meaning stories as high

meaning (see Table 6). Although mere observation inspired this feature, it's possible that those

who focus on talking about the pace of their work are experiencing job demands that exceed

resources, and are experiencing job strain, which may be negatively associated with

meaningfulness (May, Gilson, & Harter, 2004).

Atheoretical Patterns. The following words were found to be good differentiators

between high and low-meaning stories, however theoretical support for them is, at best,

suggested post-hoc.

Together. Surprisingly, the word "together" was much more associated with low-meaning

stories than high stories. I had originally investigated it figuring that it would signal

belongingness, which has been suggested many times as a pathway to meaningful work (Lips-

Wiersma & Wright, 2012; Rosso et al., 2010). But I quickly discovered that it helped predict low

meaning stories.

Upon further investigation, it seems that indeed phrases that refer to bringing *people*

together "e.g., brings people together, brings them together, brings us together" *are* typically

found in high-meaning stories. However, phrases that imply bringing *pieces* of something

together (e.g., "several departments together") are associated with *low* meaning stories. The most

predictive instantiation (see Table 7) of this was a pattern that looked only for non-people-related

uses of "together," and used those to signal low-meaningfulness.

If asked to posit a guess about this association, I suspect that focusing on things that must

be brought "together" you are in a way thinking of them as separate, which would be more

consistent with a low-construal perspective, which Morrison et al. (2016) proposed would be

related to low-meaningfulness. Note that there is general acceptance in the construal level

54

literature for the notion that this kind of focusing on separate pieces (vs. the whole) is related to low construal level. An especially interesting and relevant example of this is Burgoon, Henderson, and Markman (2013), who showcase studies that have linked abstract (high-construal) thinking to better performance on Gestalt completion tasks (i.e., in a high construal state people see 'the whole' more readily than the pieces).

Given the distinction between bringing pieces together and bringing people together, the most predictive patterns taught the algorithm to treat the two uses of 'together' as separate features.

Regex search pattern for 'people together':

```
people together|them together
```

Regex search pattern for 'pieces together':

```
[^people|them] together
```

"Have to" versus "Get to". An anecdotal observation I've made since beginning my studies in organizational psychology is that people who enjoy something, especially their jobs, tend to say things like "I *get to* do math all day" whereas people who dislike an activity feel burdened by it, and say things like "I *have to* do math all day." I have no theoretical reason for this, although it did occur to me after reading Kahn's (1990) vignettes of work engagement/disengagement. In any case, I found in this study that the phrase "have to" is a predictive signal of low-meaningfulness and "get to" is a (weaker) predictor of high-meaningfulness. I will note, also anecdotally, that the use of "get to" showed up frequently in people's separate descriptions of explicitly *why* they find their work meaningful, although these

were not used for analysis.

Summary of new language features. Table 6 and Table 7 show each pattern used in the final algorithm, grouped by whether they predict high or low meaningfulness, together with ratio scores that represent their ability to differentiate between high and low meaningfulness. A note on interpreting these ratios: In machine learning, these ratios are called 'likelihood ratios.' A likelihood ratio of "5:1" means that a feature occurs 5 times more often in one case than another. Note also that there are also "negative features" grouped under "Text does NOT contain…" headings. This is intended to indicate that the *absence* of this feature is predictive of a particular meaningfulness decision.

Table 6. High meaningfulness features. Importance of language features in predicting *high* self-reported overall work meaningfulness.

Feature	Importance Ratio
Does contain...	
Identity words	2.9:1
Work centrality words	1.4:1
Does NOT contain...	
Pace	1.4:1
Have to	1.1:0

Table 7. Low meaningfulness features. Importance of identified language features in predicting *low* self-reported overall work meaningfulness.

Feature	Importance Ratio
Does contain...	
Pace	3.9:1
Temporariness	2.5:1
Pieces together	1.2:1
Does NOT contain...	
Work centrality	1.2:1
Identity words	1.2:0
Have to	1.1:0

Goal 2: Create a Natural Language Measure of Work Meaningfulness

As a second contribution, this study introduces a new, natural-language-based measure of meaningfulness. By using a predictive model comprised of the language features outlined above, this algorithm is able to predict how meaningful people find their work from how they write about it. Specifically, this algorithm is able to predict self-reported work meaningfulness (assessed via the item "Overall, is your work very meaningful? Yes or no?") with 85% accuracy.

Cross-validation. A K-fold cross-validation test was performed, in which the text sample was divided randomly into different sub-samples so that the algorithm could be re-tested on 'different' bodies of text. This K-fold test reported that the overall accuracy (or 'skill') of my meaningfulness algorithm was a 'stable' 85%. This means that when presented with three sets of work stories that the algorithm had never seen before, it correctly predicted whether the author of those stories found their work meaningful (or not) with an average accuracy of 85%.

Relationship with collected measures. Hypothesis 5 proposed that ratings of meaningfulness produced by the natural language algorithm would correlate significantly with other measures of meaningfulness. And the related hypothesis 6 proposed that algorithm-rated meaningfulness would correlate significantly with a known outcome of meaningfulness (affective commitment). To test these hypotheses, I conducted Pearson correlation tests on the relationship between the 'probability of meaningfulness' score produced by the algorithm and other measures of meaningfulness and affective commitment. Results show that, except for human ratings of meaningfulness, these hypotheses were fully supported (see Table 8).

Human-rated meaningfulness. The algorithm's probability of meaningfulness ratings failed to show a relationship with human-ratings of meaningfulness, with a non-significant correlation of **.10**. Given the low correlations between human-rated meaningfulness and 'actual'

59

self-reported meaningfulness, it is not particularly surprising that human ratings failed to correlate with algorithm-predicted meaningfulness (which was optimized to predict self-reported meaningfulness).

Self-reported meaningfulness (single item). The algorithm's ratings showed a significant, positive correlation (.35, $p < .001$) with self-reported meaningfulness assessed with the "How meaningful do you find your current work?" item.

Self-reported meaningfulness (binary). The algorithm's ratings showed a significant, positive correlation (.31, $p < .001$) with self-reported meaningfulness assessed with the "Overall, I find my work very meaningful" item.

Work and Meaning Inventory (WAMI). The algorithm's ratings showed a significant, positive correlation (.29, $p < .001$) with self-reported meaningfulness assessed with the current most-cited Likert meaningfulness measure, Steger et al.'s (2012) Work and Meaning Inventory.

Comprehensive Meaningful Work Scale (CMWS). The algorithm's ratings showed a significant, positive correlation (.26, $p < .001$) with self-reported meaningfulness assessed with Lips-Wiersma and Wright's (2012) Comprehensive Meaningful Work Scale.

Affective commitment. The algorithm's ratings showed a significant, positive correlation (.24, $p < .001$) with self-reported affective commitment as measured by Allen and Meyer's (1990) affective commitment scale.

Table 8. Correlations algorithm-predicted probability of meaningfulness and collected measures of meaningfulness.

Hypotheses	Meaningfulness measure	Correlation with algorithm-predicted probability of meaningfulness
H5-A	Human ratings of meaningfulness	.10ns
H5-B	Self-reported meaningfulness (1-7)	.35***
H5-C	Self-reported meaningfulness (binary)	.31***
H5-D	WAMI[1]	.29***
H5-E	CMWS[2]	.26***
H6	Affective Commitment	.24***

*** $p < .001$, [1]**WAMI** = Work And Meaning Inventory (Steger et al., 2012), [2]**CWMS** = Comprehensive Work as Meaning Inventory (Lips-Wiersma & Wright, 2012).

Goal 3: Testing Construal Level & Meaningfulness

As an additional theoretical contribution of this paper, I tested the relationship between perceptions of work meaningfulness and construal level, which Morrison et al. (2016) proposed as a potential unifying mechanism to the somewhat fragmented theoretical literature surrounding work meaningfulness. As detailed above, trained raters rated each work story for the overall construal level of the language used in the story.

In Hypotheses 8 and 9, these other ratings of construal level were tested via Pearson's correlation test for a relationship with self-reported and other-rated meaningfulness.

H8: Construal level will have a significant, positive relationship with meaningfulness.

H9: Construal level will have a significant, positive relationship with Affective Commitment.

Both of these construal-related hypotheses were fully supported. By any measure, construal level seems to be positively related to perceptions of work as meaningful. When people find their work meaningful, they speak about it more in terms of its overall, zoomed-out qualities. When people find their work meaningless, they speak about it more in terms of its concrete details.

Table 9. Correlations between other-rated construal level and meaningfulness measures.

Hypotheses	Meaningfulness measure	**Other-rated Construal Level**
H8-A	Other-rated meaningfulness	.74***
H8-B	Continuous single item	.20***
H8-C	WAMI[1]	.21**
H8-D	CMWS[2]	.16**
H9	Affective Commitment	.21**

** p < .005 *** p < .001, [1]**WAMI** = Work And Meaning Inventory (Steger et al., 2012),

[2]**CWMS** = Comprehensive Work as Meaning Inventory (Lips-Wiersma & Wright, 2012).

On the validity of other-rated construal level. A note on the validity of the other-rated construal level in light of the apparent lack of validity of other-rated meaningfulness: Recall that part of the motivation for this study was the lack of definitional clarity in the meaningfulness literature. Without a clear consensus on what meaningfulness is, it was difficult to teach others to recognize it.

The literature on construal level, by contrast, is nearly in the opposite state as the meaningfulness literature: it is clearly and consistently defined, and the concept itself is objective and easy to manipulate in accordance with its definition (see Wakslak, Liberman, and Trope [2007] for a review). Additionally, many instantiations of construal level have been identified in organizational research, making it easy to generate work-related examples (Wiesenfeld, Reyt, Brockner, & Trope, 2017). All of this makes construal level much easier to recognize and teach than meaningfulness. Note also that construal level correlated significantly (.21, $p < .001$) with the convergent validity check measure (affective commitment).

This finding that construal level relates to all collected measures of work meaningfulness, as well as to affective commitment has large implications for the meaningfulness literature. Construal level may indeed be ready for further investigation as a core 'mechanism of meaning,' as Morrison et. al (2016) proposed.

Joint Relationships

In addition to the bivariate correlational results provided thus far, I also tested the incremental validity of the significant variables from related hypotheses. These variables have shown that they are significantly related to our outcomes of interest alone, but do they offer additional predictive power when combined together using a Multiple Regression (MR) model?

What best predicts whether a story *sounds* meaningful? The other-rated

meaningfulness scores provided by human raters are, effectively, ratings of how meaningful each story *sounded*. Although these ratings proved to be too unrelated to 'actual', self-reported meaningfulness to be useful to algorithm development, it may still be useful to understand what variables contribute to making somebody sound (to others) like they find their work meaningful.

Recall that the use of first-person pronouns (e.g., I, we) and action verbs (e.g., help, save, work), and positive sentiment were positively correlated with ratings of meaningfulness provided by external raters. Multiple regression analysis was used to test which of the language variables predicted other-rated meaningfulness. Results of the regression indicated that although action verbs alone ($\beta = 110.53$, $R^2 = .05$, $p < .001$) and personal pronouns alone ($\beta = 79.33$, $R^2 = .09$, $p < .001$) provide some predictive power, the significance of action verbs falls away when the two are combined. Personal pronouns, however, still remain significant in the combined model ($\beta = 68.25$, $p < .01$), which predicts only as well as personal pronouns alone ($R^2 = .09$).

When positive sentiment is added to the model, the significance of both action verbs and pronouns falls away, with positive sentiment showing the only significant slope ($\beta = 16.58$, $p < .001$). Ultimately, a model excluding action verbs and combining personal pronouns and positive sentiment accounted for the most variance. In this model, both positive sentiment ($\beta = 16.62$, $p < .001$) and personal pronouns ($\beta = 39.16$, $p < .05$) show significant incremental validity, and the full model accounts for ($R^2 = .29$) of the variance in other-reported meaningfulness.

Summary. In sum, the best predictor of whether a person *sounds like* they find their work meaningful (to others) is whether they talk about it with a positive tone of (written) voice, and use lots of personal pronouns (I, me, you).

What best predicts self-reported meaningfulness? While positive sentiment and personal pronouns together accounted for 29% of the variance in other-rated meaningfulness, the

same model only accounted for 5% of the variance in 'actual' *self*-reported meaningfulness (rated on a continuous scale from 1 to 7), and only positive sentiment remained significant (β = 9.99, p < .001). So then, which of the variables collected 'traditionally' (outside the machine learning algorithm) best predicts self-reported meaningfulness?

I tested several combinations of language features (i.e., positive sentiment, abstract word score, construal level, identity words, personal pronouns, and action verbs). The best model explained (R^2 = .11) of the variance in self-reported meaningfulness by combining positive sentiment (β = 8.80, p < .001) with identity words (β = 0.21, p < .001). Based on this result, it seems that the best indicators of self-reported meaningfulness are whether somebody writes about work with a positive tone and also starts their work story with an identity statement (e.g., "I am a…").

Algorithm-predicted meaning. As a final consideration, note that a regression model consisting only of the "probability of meaning" score produced by this study's NLP algorithm explained more of the variance in self-reported meaningfulness than did the best model of independent language features (β = 4.29, p < .001, R^2 = .15). When positive sentiment was added to this model (β = 6.83, p < .01), variance explained improved to R^2 = .18, suggesting that a future version of the algorithm might benefit from incorporating positive sentiment into its probability estimates.

DISCUSSION

Contribution 1: Language Reveals Meaningfulness

The core assertion of this study, that people's sense of meaningfulness comes through in how they talk about their work, seems to hold true. There *are* linguistic indicators of work meaningfulness, and it's possible to use such indicators to predict from speech alone whether a person finds their work meaningful.

Validation of Podolny et al.'s (2004) theory. The study also served as a partial test of Pdonlony et al.'s (2004) study on linguistic indicators of meaningful work. And results suggest that Podolny et al.'s (2004) conclusions have merit: first-person pronouns and action verbs (e.g., "I file TPS reports" vs. "TPS reports are filed") do seem to have a role in creating *the impression of* meaningful work. This finding itself may have implications about what people think the term 'meaningfulness' means *for others*.

Theoretical implications. Given the new language features found to aid in the prediction of self-reported meaningfulness, it seems like identity ("I am…") and work centrality ("worry/care") are worthy of further investigation for their relationship with feelings of meaningfulness in work. Likewise, notions of job strain ("pace") and a sense of burden ("have to") may deserve further investigation for their relationship with feelings of meaning*less*ness in work. The finding that words emphasizing separate pieces are associated with a sense of low-meaningfulness seems to provide further support for the proposition in this study that a low-construal of work (seeing it for its separate, concrete details) is associated with low meaningfulness.

Future directions. The language features identified in this study are just a beginning. They are by no means comprehensive. There are likely myriad other linguistic signals of meaningfulness left to be discovered. In this section, I will suggest some promising potential avenues that could lead to new and perhaps even more-predictive linguistic signals.

Further mine Podolny et al. (2004). Podolny et al. (2004) proposed other operationalizations of their notion of self-distance from work being associated with meaningfulness. For example, the proposed that "you-language" phrases like "You're expected to" signal a sense of self-distance from work. Although these were not included in my initial hypotheses, I did run some preliminary (word frequency) tests on a couple of them to test for their ability to discriminate between high and low meaning. On the first attempt, they were not powerful enough on their own to significantly contribute to the algorithm's predictive accuracy. Anecdotally however, I can attest from reading the work stories that some of the additional language signals of low/high-meaning provided by Podolny et al. (2004) do indeed seem accurate for extreme examples (very high or low meaning on the continuous scales), but may not be as useful at differentiating middle-range scores, which may be why they failed to help predict overall in my preliminary tests.

Additionally, Podolny et al. (2004) noted in their discussion that their proposed linguistic indicators of meaningfulness were likely too appear together. This could explain the high correlation found in this study between personal pronouns and action verbs. It could also suggest that when testing any additional features suggested by Podolny et al. (2004), one should be rigorous about checking for incremental predictive validity over and above the other features.

Ultimately, I think that there is much more to be mined from Podolny et al.'s (2004) paper than was tested in this study, and would like to see future tests of their suggestions use

even more complex text processing techniques (e.g., custom decision trees and sentence-level analysis) to fully test all of their propositions.

Limitations. In my introduction I gave a brief tour of some of the problems facing the literature on work meaningfulness. Namely, there are many different definitions of meaningfulness, and no consensus yet on what the 'best' definition is. Another limitation I have not yet mentioned is that meaningfulness is discussed near-universally as an overall evaluation of one's work, even though it is possible that the experience of finding your work meaningful may also occur in-the-moment, rather than retrospectively. It is possible that capturing meaningfulness as it occurs during the work day, versus as an overall judgement, may increase the strength of the 'signal' and allow for more accurate assessment of the construct.

Contribution 2: A Natural Language Measure of Meaningfulness

There is no shortage of Likert-based measures of meaningful work. Bailey et al. (2016) list 25 different measures. To my knowledge, this study introduces the first measure of work meaningfulness that utilizes language patterns rather than Likert items. However, the NLP measure at the moment is only validated on samples of text where people were talking about their work in response to a certain, generic prompt. It is unknown how far it could generalize to other samples of text. There are many instances where similar samples of text may be generated (e.g., job analysis, talking about past jobs in job interviews).

Distributing the measure. Likert-based measures are designed for paper-and-pencil testing, and thus can be distributed easily via PDF file or within an article. However, there is little precedent for distributing natural language-based measures of psychological constructs. In the hopes that this measure could be helpful to other researchers and practitioners, the code for this measure will be uploaded (along with instructions for use) to GitHub as soon as a paper

introducing it is published. The code will be available under a Creative Commons - BY (CC-BY) license, which means it will be free for modification, distribution, and use with proper credit.

Future directions. Ultimately, with further refinement and testing on new samples of text, I hope that this measure could one day be used on open-ended survey responses and more passively-collected natural language data, like emails and customer service phone call transcripts. It may also be interesting to translate the measure for use with other languages.

In terms of accuracy improvement, I think that a hybrid approach that combines the two approaches in this study (XGBoost/bag-of-words and Naive Bayes/binary discrimination), along with a custom decision tree classifier (which would allow me to say "If you see this pattern, it's always low meaningfulness"), could yield higher prediction accuracy, perhaps nearing 100%.

Limitations. It should be noted that machine-learning algorithms like those in this study utilize multiple regressions to achieve their predictive results. Running so many regressions can threaten generalizability by inflating type 1 error. That is, there's a risk that so much trial and error can result in finding 'successful predictors' that may just be idiosyncratic to this data. The K-Fold test performed in this study is the standard practice for addressing this concern. This K-Fold method put the generalizability of the algorithm to the test by checking its predictive validity on multiple samples of data that it hadn't seen before, and it passed these tests by maintaining its initial accuracy of 85% on new data. However, it should be noted that one large generalizability concern still remains: That the data used to train and test the algorithm is highly structured and may be too different from text data found organically in the 'real world'.

Contribution 3: Construal Level is Related to Meaningfulness

Construal level shows *at least* a medium-sized relationship with several different measures of work meaningfulness. I suspect with further training on recognizing examples of

high/low construal level, raters could capture more of the variance in construal level, and that refined rating may relate even more strongly to meaningfulness than this initial attempt.

This construal level relationship has many implications for research on meaningful work. First, it provides evidence to suggest that construal level may deserve a place as a central feature in definitions of meaningfulness, and further development along Morrison, Walker, and DeShon's (2016) theoretical line seems warranted.

Second, it suggests that construal level may indeed be a much-needed unifying component in work meaningfulness theory and should be included in discussions of the nature of experienced work meaningfulness. A high construal level perspective may be a precursor to perceiving meaning through *any* specific pathway to meaningfulness like those identified by Lips-Wiersma and Wright (2012). This construal level finding should be treated as the beginning ("tip of the iceberg") of what should be a much larger investigation of construal level's relationship with work meaningfulness.

The relationship between construal level and work meaningfulness may also have wide implications for the meaning-making literature. Meaning-making refers to the process through which individuals are able to create a sense of meaning in their work (Frankl, 1962; Rosso et al., 2010). It's possible that 'raising construal level', for instance by articulating a broad vision, is the route through which meaning-making travels.

Future directions. This study was designed primarily to test the relationship between meaningfulness and language. It included a test of construal level as a secondary, opportunistic contribution. However, a more direct test of construal level and meaningfulness could shed even further light on the relationship.

A future study might first manipulate construal level, and then ask about the

meaningfulness of the participant's work. If construal level of work and meaningfulness of work are related as Morrison et al. (2016) proposed and the results of this study suggest, then a manipulated construal level could spill over onto perceptions of meaningfulness. Specifically, meaningfulness ratings should be higher in a high-construal condition than in a low-construal condition.

Contribution 4: The Work Stories Corpus

Studs Terkel's (1972) book *Working* and Bowe et al.'s (2000) book *Gig* are both 'just' collections of assorted people talking about their work. With 1754 citations for the former, and 90 for the latter, each has been highly generative for researchers seeking to understand work experiences, and in some cases (i.e., Podolny et al. [2004]) to study natural language use. There are two major limitations hindering the more widespread use of these books in research: First, it is hard to find digital copies of the books for easy analysis. Second, neither book (obviously) includes any explicit psychological data on the authors of its stories, which limits their value in looking for language patterns that correlate with various psychological phenomenon (i.e., there are no criteria to compare to, requiring manual coding).

The present study created a large, digital collection of Work Stories similar to those in Terkel's (1972) *Working* and Bowe et al.'s (2000) *Gig* and paired them with a great deal of psychological data about the authors of each story. My hope is that this collection of work stories and accompanying data will empower other researchers to discover new qualitative insights about work and to develop new work-related natural language measures. To maximize the impact and generativity of this work story dataset, I plan to post it free and Open Access on the Open Science Framework, following the publication of a 'data paper' which will explain the dataset and give researchers something to cite, so that this project can be credited when the data

is used.

PRACTICAL IMPLICATIONS

Watch for Language Cues of Meaningfulness

If nothing else, I hope this paper has taught you a cool party trick: Next time you're at a party, and you ask someone "What do you do?", listen to the first words out of their mouth. If they start with an identity statement like "I am a..." instead of "I work at..." or "Currently, I…", and they say it with a positive tone of voice, there's a chance they find their work meaningful. Similarly, if you hear them lamenting the pace of their work or all the separate pieces that "have to" be brought together, it may signal that they find their work low in meaning.

Practitioners could be trained to look for these language signals of low and high meaning whenever people talk about their job (e.g., in interviews, job analyses, and performance feedback meetings).

A New Tool for Practitioners

Practitioners will be able to download and use the NLP meaningfulness algorithm created in this study freely as soon as it is published on GitHub. I recommend trying it especially in contexts where a person is describing their job, such as describing past jobs in job interview transcripts or describing their current job in open-ended survey responses. Modifications and improvements to the measure can also be made easily by 'forking' the GitHub repository.

Watch for Construal Level Fluctuations When People Talk About Their Work

The construal level findings in this study have much larger implications for theory than practice. However, it could still be useful for practitioners and managers to know that if they hear a worker describing their job in too many concrete details (e.g., "TPS reports, calendar,

bathroom, desk") it may be a signal that they're experiencing some meaninglessness. If they

describe it in abstract terms and talk about the broader nature of their work (e.g., "keeping

everybody up to speed" instead of "filing reports"), it may signal that they find their work highly

meaningful.

APPENDIX

Training Procedure for Construal Level

In the following sections, I will outline the procedure used to train the undergraduate raters to recognize construal level in the work stories.

Reading Assignment. Each member of the rating team read Trope and Liberman's seminal (2010) paper examining construal level in detail. This paper includes several examples of construal level's antecedents and consequences, and explains the definition of the construct at length.

Response paragraph. Each rater was asked to send a short response paragraph summarizing their impression of the definition of construal level. These paragraphs were all deemed to indicate satisfactory basic understanding.

Guided examples. As a group, raters were shown 3 example stories from pilot data. Two examples of clearly high-construal stories (lots of abstract connections and language), and one example of a clearly low-construal story (which mostly discussed the concrete features of the work). Of the two high-construal examples, one was a high-meaning/high-construal story, and one was a low-meaning/low-construal story. In advance of the training session, I marked-up the example stories, highlighting language signals that struck me (as a subject matter expert) as particularly high or low construal. I discussed each of these highlighted sentences, and explained why I thought they represented high/low construal.

Training Procedure for Meaningfulness

In the following sections, I will outline the procedure used to train the undergraduate raters to recognize meaningfulness in the work stories.

Reading assignment. Each undergraduate rater read the Morrison, Walker, and DeShon

(2016) definition of meaningfulness paper as an introduction to the construct of meaningfulness. It should be noted that this paper includes a focus on construal level as the potential unifying mechanism of meaningfulness, which could well have alerted the raters to the hypotheses being tested, and could have explained the strong correlation between other-rated and self-reported meaningfulness.

Writing assignment. Each rater wrote two, 500-word essays (of the same length as those written by participants). In one essay, they were asked to write about the most meaningful work they ever participated in. In the second essay, they were asked to write about the most meaningless work they ever participated in. In both essays, they were asked to include an additional paragraph introspecting about how they talked about each work experience. The idea here was to get them thinking about language signals that they used when talking about meaningful/meaningless work to help them recognize such signals in the stories of others.

Guided examples. Using the same example stories used to train construal level, raters were walked through sections in each work story that indicated to me (as a subject matter expert) that the person found their work meaningful. I also explained the logic behind the signals I pointed out, noting how each represented a connection (or lack thereof) to the author's values, and/or represented a common pathway to meaningfulness.

REFERENCES

REFERENCES

Allan, B. A., Duffy, R. D., & Collisson, B. (2018). Helping others increases meaningful work: Evidence from three experiments. *Journal of Counseling Psychology, 65*(2), 155-165.

Arnold, K., Turner, N., Barling, J., Kelloway, E. K., & McKee, M. C. (2007). Transformational leadership and psychological well-being: The mediating role of meaningful work. *Journal of Occupational Health Psychology,12,* 193-203.

Bailey, K., Yeoman, R., Madden, A., Thompson, M., Kerridge, G. (2016). A Narrative Evidence Synthesis of Meaningful Work: Progress and Research Agenda. A paper presented at the meeting of the Academy of Management, Anaheim, CA.

Barrick, M. R., Mount, M. K., & Li, N. (2012). The theory of purposeful work behavior: The role of personality, higher-order goals, and job characteristics. *Academy of Management Review, 38*(1), 132–153. https://doi.org/10.5465/amr.2010.0479

Britt, T.W., Dickinson, J.M., Castro, C.A. & Adler, A.B. (2007). Correlates and consequences of morale versus depression under stressful conditions. *Journal of Occupational Health Psychology, 12*, 34-47.

Bowe, J., Bowe, M., & Streeter, S. C. (2000). Gig: Americans talk about their jobs at the turn of the millennium. New York: Crown Publishers.

Brownlee, J. (2018, May 23). A Gentle Introduction to k-fold Cross-Validation. Retrieved from https://machinelearningmastery.com/k-fold-cross-validation/

Brysbaert, M., Warriner, A.B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*, 904-911.

Bunderson, J. S., & Thompson, J. A. (2009). The call of the wild: Zookeepers, callings, and the double-edged sword of deeply meaningful work. *Administrative Science Quarterly, 54*(1), 32-57.

Burgoon, E. M., Henderson, M. D., & Markman, a. B. (2013). There are many ways to see the forest for the trees: A tour guide for abstraction. *Perspectives on Psychological Science, 8*, 501–520. https://doi.org/10.1177/1745691613497964

Chalofsky, N. E. (2010). Meaningful workplaces: Reframing how and where we work. San Francisco, CA: John Wiley & Sons.

Champoux, J. E. (1992). A multivariate analysis of curvilinear relationships among job scope, work context satisfactions, and affective outcomes. *Human Relations, 45*(1), 87-111.

Cohen-Meitar, R., Carmeli, A., & Waldman, D. A. (2009). Linking meaningfulness in the workplace to employee creativity. The intervening role of organizational identification and positive psychological experiences. *Creativity Research Journal, 21*, 361-375.

Diesner, J., Frantz, T. L., & Carley, K. M. (2005). Communication networks from the Enron email corpus "It's always about the people. Enron is no different". *Computational & Mathematical Organization Theory, 11*(3), 201-228.

Douglas, K., & Carless, D. (2009). Abandoning the performance narrative: Two women's stories of transition from professional sport. *Journal of Applied Sport Psychology, 21*(2), 213-230.

Duffy, R. D., Allan, B. A., Autin, K. L., & Bott, E. M. (2013). Calling and life satisfaction: It's not about having it, it's about living it. *Journal of Counseling Psychology, 60*, 42-52.

Fairlie, P. (2011). Meaningful work, employee engagement, and other key outcomes: Implications for human resource development. *Advances in Human Resources, 13*, 508-525.

Figueroa, R. L., Zeng-Treitler, Q., Kandula, S., & Ngo, L. H. (2012). Predicting sample size required for classification performance. *BMC medical informatics and decision making, 12*(8), 1-10.

Förster, J., Friedman, R. S., & Liberman, N. (2004). Temporal construal effects on abstract and concrete thinking: Consequences for insight and creative cognition. *Journal of Personality and Social Psychology, 87*(2), 177-189.

Frankl, V. E. (1962). Man's search for meaning. New York: Simon and Schuster.

Glazer, S., Kozusznik, M. W., Meyers, J. H., & Ganai, O. (2014). Meaningfulness as a resource to mitigate work stress. In S. Leka, & R. R. Sinclair (Eds.), *Contemporary occupational health psychology: Global perspectives on research and practice* (Vol. 3, pp. 114-130). Chichester, UK: Wiley-Blackwell.

Grant, A. M. (2008). The significance of task significance: Job performance effects, relational mechanisms, and boundary conditions. *Journal of Applied Psychology, 93*(1), 108-124.

Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. Organizational Behavior and Human Performance, 16(2), 250-279.

James, W. (1985). The varieties of religious experience: A study in human nature. [Powerpoint slides] Retrieved from http://93beast.fea.st/files/section1/James%20-%20Varieties%20of%20Religious%20Experience.pdf

Jiang, L., & Johnson, M. J. (2018). Meaningful work and affective commitment: A moderated mediation model of positive work reflection and work centrality. *Journal of Business and*

Psychology, 33(4), 545–558. https://doi.org/10.1007/s10869-017-9509-6

Lips-Wiersma, M., & Wright, S. (2012). Measuring the meaning of meaningful work: Development and validation of the Comprehensive Meaningful Work Scale (CMWS). *Group & Organization Management, 37*(5), 655–685. https://doi.org/10.1177/1059601112461578

Lepisto, D. A., & Pratt, M. G. (2016). Meaningful work as realization and justification: Toward a dual conceptualization. *Organizational Psychology Review, 7*, 99-121. https://doi.org/10.1177/2041386616630039

Kahn, J. H., Tobin, R. M., Massey, A. E., Anderson, J. A., The, S., Journal, A., & Summer, N. (2016). Measuring emotional expression with the Linguistic Inquiry and Word Count, *The American Journal of Psychology, 120*(2), 263–286.

Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal, 33*(4), 692-724.

K-Fold Cross-Validation. (2018). In Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Cross-validation_(statistics)#k-fold_cross-validation

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50*(4), 370-396.

Maslow, A. (1969). The farther reaches of human nature. Journal of Transpersonal Psychology, 1(1), 1–9

Mahmud, J. (2015). IBM Watson Personality Insights: The science behind the service. Retrieved from https://developer.ibm.com/watson/blog/2015/03/23/ibm-watson-personality-insights- behind-service/

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research, 30*, 457–500. https://doi.org/10.1613/jair.2349

Malsburg, T. V. (n.d.). How to correctly calculate worker compensation for Amazon Mechanical Turk. Retrieved March 26, 2017, from https://tmalsburg.github.io/blog/how-to-correctly-calculate-worker-compensation-for-amazon-mechanical-turk/

Martela, F., & Steger, M. F. (2016). The three meanings of meaning in life: Distinguishing coherence, purpose, and significance. *The Journal of Positive Psychology, 11*(5), 531-545. https://doi.org/10.1080/17439760.2015.1137623

May, D. R., Gilson, R. L., & Harter, L. M. (2004). The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work. *Journal of Occupational and Organizational Psychology, 77*(1), 11-37.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*(1), 81-90.

Morrison, MA., (2016). Increasing the meaningfulness of work with motivational self-transcendence. Paper presented at the meeting of the Academy of Management, Anaheim, CA.

Morrison, MA., Walker, R, DeShon, R. (2016). Toward a comprehensive definition of work meaningfulness, Paper presented at the meeting of the Society for Industrial-Organizational Psychology, Orlando, FL.

Naive Bayes classifier. (2018, August 02). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Naive_Bayes_classifier

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender Differences in Language Use: An Analysis of 14,000 Text Samples. *Discourse Processes, 45*(3), 211–236. https://doi.org/10.1080/01638530802073712

Pennebaker, J.W., Booth, R.J., Boyd, R.L., & Francis, M.E. (2015). Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).

Podolny, J. M., Khurana, R., & Hill-Popper, M. (2004). Revisiting the meaning of leadership. In B. M. Staw & R. M. Kramer (Eds.), *Research in Organizational Behavior, 26*, 1–37.

Rothmann, S., & Hamukang'andu, L. (2013). Callings, work role fit, psychological meaningfulness and work engagement among teachers in Zambia. *South African Journal of Education, 33*(2), 1–16.

Rosenberg, A., & Hirschberg, J. (2005). Acoustic/prosodic and lexical correlates of charismatic speech. Paper presented in Ninth European Conference on Speech Communication and Technology.

Rosso, B. D., Dekas, K. H., & Wrzesniewski, A. (2010). On the meaning of work: A theoretical integration and review. In A. P. Brief & B. M. Staw (Eds), *Research in Organizational Behavior, 30*, 91–127. https://doi.org/10.1016/j.riob.2010.09.001

Rosso, B. D., Dekas, K. H., & Wrzesniewski, A. (2011). Corrigendum to "On the meaning of work: A theoretical integration and review". In A. P. Brief & B. M. Staw (Eds.), Research in Organizational Behavior, 31, 277.

Santorini, B. (1990). Part-of-speech tagging guidelines for the Penn Treebank Project (3rd revision). Technical Reports (CIS), 570.

Schnell, T., Höge, T., & Pollet, E. (2013). Predicting meaning in work: Theory, data, implications. *The Journal of Positive Psychology, 8*, 543-554.

Schwartz, A., Eyal, T., & Tamir, M. (2018). Emotions and the big picture: The effects of construal level on emotional preferences. *Journal of Experimental Social Psychology, 78,* 55–65. https://doi.org/10.1016/j.jesp.2018.05.005

Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues, 50*(4), 19-45.

Steger, M. F., Dik, B. J., & Duffy, R. D. (2012). Measuring meaningful work: The work and meaning inventory (WAMI). *Journal of Career Assessment, 20*(3), 322-337.

Stillman, T. F., Baumeister, R. F., Lambert, N. M., Crescioni, A. W., DeWall, C. N., & Fincham, F. D. (2009). Alone and without purpose: Life loses meaning following social exclusion. *Journal of Experimental Social Psychology, 45*(4), 686–694. https://doi.org/10.1016/j.jesp.2009.03.007

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. https://doi.org/10.1177/0261927X09351676

Terkel, S. (1972). Working: People talk about what they do all day and how they feel about what they do. Bridgewater, NJ: The New Press.

Thompson, K. (1968). Programming techniques: Regular expression search algorithm. *Communications of the ACM, 11*(6), 419-422.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*(2), 440

Updegraff, J. A., Emanuel, A. S., Suh, E. M., & Gallagher, K. M. (2009). Sheltering the self from the storm: Self-construal abstractness and the stability of self-esteem. *Personality and Social Psychology Bulletin, 36,* 97–108.

Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social psychology, 57*(4), 660.

Van der Cruyssen, L., Heleven, E., Ma, N., Vandekerckhove, M., & Van Overwalle, F. (2014). Distinct neural correlates of social categories and personality traits. *NeuroImage, 104,* 336–346. https://doi.org/10.1016/j.neuroimage.2014.09.022

Vauclair, C. M., Hanke, K., Fischer, R., & Fontaine, J. (2011). The structure of human values at the culture level: A meta-analytical replication of Schwartz's value orientations using the Rokeach Value Survey. *Journal of Cross-Cultural Psychology, 42*(2), 186-205.

Wakslak, C., Liberman, N., & Trope, Y. (2007). Construal levels and psychological distance: Effects on representation, prediction, evaluation, and behavior. *Journal of Consumer Psychology, 17*(2), 83–95. https://doi.org/10.1016/S1057-7408(07)70013-X.Construal

Weiss, H. M., & Rupp, D. E. (2011). Experiencing work: An essay on a person-centric work psychology. *Industrial and Organizational Psychology, 4*(1), 83-97.

Wiesenfeld, B. M., Reyt, J.-N., Brockner, J., & Trope, Y. (2017). Construal level theory in organizational research. *Annual Review of Organizational Psychology and Organizational Behavior, 4*(1), 367–400. https://doi.org/10.1146/annurev-orgpsych-032516-113115