



Norwegian
Business School

This file was downloaded from BI Open, the institutional repository (open access) at BI Norwegian Business School <https://biopen.bi.no>

"This is an Accepted Manuscript of a book chapter published by Springer in Respect for Thought. Theory and History in the Human and Social Sciences from 2020 available online: https://doi.org/10.1007/978-3-030-43066-5_17"

Copyright policy of Springer, the publisher of this chapter.

Authors whose work is accepted for publication in a non-open access Springer or Palgrave Macmillan book are permitted to self-archive the accepted manuscript (AM), on their own personal website and/or in their funder or institutional repositories, for public release after an 24 month embargo period

<https://www.springernature.com/gp/open-research/policies/book-policies>

WITTGENSTEIN'S REVENGE: HOW SEMANTIC ALGORITHMS CAN HELP SURVEY RESEARCH ESCAPE SMEDSLUND'S LABYRINTH.

Jan Ketil Arnulf, BI Norwegian Business School

INTRODUCTION

Over more than three decades now, Jan Smedslund has been publishing a series of studies on pseudo-empiricity (Smedslund, 1978, 1987, 1988, 1994, 1995, 2012, 2015). Through his analysis of psychological experiments and measurement instruments, he has shown how we are already in possession of the knowledge that the studies claim to uncover. Taken literally, the actual studies should be superfluous as they tell us nothing new. It seems justified to say that his criticism has gone largely unheeded by the research community. While his ideas have generated some debate, the wider research community does not seem to bother and research practice therefore does not seem to change. References to the problem of logical and semantic structures in research remain hard to publish, keeping psychology trapped in what I henceforth will refer to as “Smedslund’s labyrinth”: re-discovering what we already know through research designs that merely illustrate what is reasonable.

The purpose of the present chapter is to look at Smedslund’s description of pseudo-empiricity and test some of his central claims using computing science. I will show that some of Smedslund’s ideas are compatible with general principles of computing science as embedded in programming languages and high-level algorithms, and that they share common roots. Computing science used in conjunction with psychology might therefore offer a possible way out of Smedslund’s labyrinth.

If we can use computing algorithms to prove some of Smedslund’s ideas experimentally, we can possibly also turn this research agenda into a true psychological endeavor: Why are his warnings so difficult to grasp, even to highly trained researchers? If Smedslund is right, why do we not know what we already know? If the pseudo-empirical studies only explore what is given in the research questions, why are we so unskilled at meta-linguistic inferences about knowledge? Therefore, it may be justified to propose that if we can simulate Jan Smedslund’s claims about Psycho-logic (PL) (Smedslund, 1995, 2012), we can make our subjective blindness the object of psychological research, paradoxical though it may seem. Through their roots in philosophy and formal logics, some assertions of PL could be demonstrated through the use of computer algorithms. We can actually show empirically that prevalent practices in psychometric research produce data that are predictable a priori. To put it bluntly, we can to some extent know what people will answer in Likert-scale surveys prior to obtaining their answers.

At the moment of writing, this type of research has been documented in a number of publications (Arnulf & Larsen, 2015; Arnulf, Larsen, Martinsen, & Bong, 2014; Arnulf, Larsen, Martinsen, & Egeland, 2018; Arnulf, Larsen, & Martinsen, 2018; Gefen & Larsen, 2017; Nimon, Shuck, & Zigarmi, 2015), but is still widely unrecognized by the scientific community. There are probably two main reasons for this. The first is that methodological paradigms in science tend to perpetuate themselves through publication practices (van Schuur & Kiers, 1994). The second reason is more psychologically interesting: The sometimes amazing cognitive capabilities of the human brain are also affected by restrictions that make us error-prone and blind to shortcomings. We find it hard to believe statements that are counter-intuitive and require cognitive efforts in understanding (Kahneman, 2011; March & Simon, 1958; Todd & Gigerenzer, 2003). For decades now, we have used computers to overcome our more obvious shortcomings in memory and calculating power. Further progress in analytical techniques may help us overcome even more advanced types of restrictions. Computers can simulate our cognitive structures and make us aware of what we know by implication of what we already know.

This is where I think psychology may even escape some of Smedslund's most dire predictions by accepting the truth of his theory. When he claims that "psychology can never be an empirical science" (Smedslund, 2016), there is now a new twist to this: We may overcome this problem by exposing our cognitive shortcomings through digital algorithms. By exploring the borderline between logical and empirical problems using digital tools, we may actually push philosophy back a few steps and make our own mental restrictions accessible to empirical research. Recent research on cognitive systems (Dennett, 2012) now emphasize the distinction between competence (what the system can achieve) and comprehension (what the system can explain about itself). By exploring the difference between competence and comprehension (the performance of our linguistic capabilities and our understanding of it) we may find answers to why it is so difficult to know what we already know.

The present chapter will first display some existing empirical findings that support the claims above. These findings sometimes seem confusing to people and require some detailed theoretical explanations. To capture and keep the readers' interest however, this chapter will begin with the findings so far, and work its way backwards through the explanations. Along the way, contributions from various traditions and ages will be presented. In the final paragraphs, I will try to integrate some of the viewpoints of the various theoretical explanations offered, and also outline a possible agenda for future research.

DIGITAL ALGORITHMS IN PSYCHOLOGY, STATUS 2017

In 2014 I thought I had discovered a disturbing finding for research using Likert-Scale surveys. Together with my co-authors Kai Larsen, Øyvind Martinsen and Chi How Bong, we published a study in the peer-reviewed journal PLoS One showing how more than 86% of the variation in the statistics from survey responses was predictable a priori (Arnulf et al., 2014). I was excited and thought that others would be, too. While I did not

think people would readily embrace the method itself, I hoped for a surprised recognition of the simple fact that the findings of a major research paradigm were obtainable in advance. There were a few initial reactions, but the scientific community has so far been silent, even as the findings have been corroborated in independent studies (Gefen & Larsen, 2017; Nimon et al., 2015).

The study analyzed data from some of the most commonly used survey instruments in organizational psychology. In this field, there are literally hundreds or even thousands of studies that explore leadership and motivation with the survey instruments we used, such as the MLQ (Avolio, Bass, & Jung, 1995), the LBDQ (Stogdill, 1963), or scales measuring various types of motivations (Kuvaas, 2006). These instruments have for many years been the gold standard of “measurement” in this research area, a prerequisite to publish in high-ranked journals (Bagozzi, 2011; Michell, 2013; Yukl, 2012). The respondents comprised four big samples from different organizations, making sure that the findings were no coincidence.

While the exact mechanisms will be explained in more detail below, I will here explain in a simple way what we found: Surveys used in organizational psychology usually explore relationships between “constructs” such as different types of leadership, different types of motivation, and their effects on work processes in organization. The researchers typically want to know if one type of leadership is more effective than another, and which psychological processes are involved in producing these effects. A typical research design may therefore imply asking participants in organizations about their perceptions of their managers, how they think about themselves and their motivations, and about the quality and intensity of the work they are doing.

The usual way to analyze these data is to make statistical explorations of the way that these answers are linked together, using correlations, regression equations or complex structural equation models (SEM) that render quantitative descriptions of how the constructs are linked together (Bagozzi, 2011; Jöreskog, 1993; MacKenzie, Podsakoff, & Podsakoff, 2011; Podsakoff, MacKenzie, & Podsakoff, 2012).

By contrast, our semantics project begins by using only the questions from the survey questionnaires themselves, their “items” as they are termed. We feed them as input to digital semantic algorithms that can estimate to which degree these sentences have overlapping meaning. Such algorithms will usually give a number between 0 and 1.0 that indicates if the sentences share meaning in their content.

We use the numbers we get from the algorithms to predict or “guess” what the correlations between the survey items would be. The results were beyond my initial expectations. Depending on the assumptions, we could predict the correlations between leadership, motivation, and the outcomes in the surveys quite accurately. The semantic values captured in the best case 86% of the variation in correlations, but more importantly: in regression equations, this level of explanation was enough to predict the actual correlations as they were created by human respondents down to two decimals.

I remember showing the tables of correlations to the British professor in organizational psychology Adrian Furnham. He looked puzzled at it for a moment, then asked: “But if the numbers simply support what we already found, isn’t that just a confirmation of our original results?” “Yes, in a way,” I replied, “but if we could

obtain the numbers simply by running the questions through a machine, we wouldn't need to ask people, would we?" I could see him reflect for a moment, then nod. "Quite," he agreed.

Working with the findings throughout the analytical process, I had constantly sought someone to prove me wrong. The numbers were simply too good, and I was constantly expecting that someone could point to a flaw in the arguments, showing that the match between semantics and survey statistics would be an artifact or a product of a mistaken sort of analysis. That person never appeared.

Instead, I met a number of researchers who kept reminding me of Jan Smedslund. Most of them would be his previous colleagues or students. Whenever I called on a statistician, a methodologist or a psychological researcher, they would chew on my findings for a while, not coming up with a better explanation, and then shrug: "It reminds me of some of Jan Smedslund's stuff, the sort of ideas he has always been talking about. Maybe you should ask him."

I will return to the relevance of Smedslund's ideas in later sections, but first a few words about the reviews that we got as the first article made its round in attempts at publication. As the article was reviewed in journals addressing organizational psychology, the reviewers generally omitted mentioning the fact that commonly reported findings could be reproduced without empirical materials. For all their methodological sophistication, this fact seemed to be the unmentionable elephant in the room. Instead, they usually recommended a rejection of the article because of its unconventional method of using digital text algorithms.

I want to quote three reviewers as their viewpoints shed interesting light about why we do not know what we already know, the a priori truths in pseudo-empirical research. One reviewer stated openly that he had no idea what "semantic algorithms" were, and so he had Googled it. What he had read on Google, he said, was unconvincing to him, and so he suggested that the manuscript should be rejected. I replied to the editor that the reviewer paradoxically had been using a text algorithm (Google search) to investigate text algorithms, leading him to declare a disbelief in text algorithms because of what he found with through the use of one (the editor agreed, and asked me to re-submit, but to wait until he himself had quit his post).

Another reviewer made a better and more informed attempt, which we have met over and over again: Maybe we were misinterpreting the findings when we claimed that they contested the empirical research. Maybe the replication of the data structures instead supported their truthfulness. In other words, we had just found what research has already established, and so it wasn't the research findings, but instead it was *our* research that was superfluous and did not deserve to be published.

Yet a third reviewer added that the text algorithms probably only reflected what people know because the research findings had been disseminated. In other words, we had used language research to find that people had already adopted the findings from leadership research.

Unawares, these three reviewers were articulating an explanation, not for our results, but for why we struggle to understand what we already know. This is a meta-linguistic phenomenon called "competence without

comprehension”, that we know how to use the language without knowing exactly how it works. I will return to this phenomenon in a later section.

First of all, the reviewer who googled the algorithms seemed to take computerized tools for granted without reflecting on what they really do. Computers are machines that apply the calculating powers of language, known as formal logics, to derive answers we are looking for from what we already know. It is sometimes hard for people to understand this, but formal logics is by its nature truth-preserving. Logical processes can strictly speaking not create new. A computer can only draw conclusions from the information already available to it. Often, we draw on this information because it is accumulated by others and so is new to the user. But most of the time, we let ourselves be amazed by how the computer is thinking in a different way from humans, more systematically and more stringently. The computer works by systematically exploiting what it already knows. One may compare it to a thinking phone book. In my younger years, possessing a phone directory, I “knew” all the phone numbers in Oslo in the sense that they were in my possession. I still had to look them up, at the risk of not finding the number I was looking for. If programmed correctly, a computer will arrive at the right number through rigid application of the same procedure, proving that it always knows what it already knows.

The second reviewer’s reply revealed that he judged our findings to be valid replications of empirical research, but that he was obviously indifferent to whether knowledge is derived from empirical methods or logical deductions. A bit curious for a trained researcher, it still reflects a long-lasting controversy between rationalists and empiricists in philosophy. Whatever one’s position on this debate, it testifies to the fact that humans are just as surprised to learn what is logically derived as what is empirically detected. We seem to want or need the information precisely because it isn’t obvious to us. We do not care how it was derived as long as there is some validity to it. At this point, reviewer 2 was voicing a version of scientific psychology that Jan Smedslund has been fighting for years. It is a discipline that at great cost goes to great lengths to tell us what we already know, what Jon Elster (2011) has called “hard obscurantism” and a waste of time and efforts in science. The a priori given answer is provided through a method so opaque to most people that they are barred from disputing it.

The third reviewer’s comment is more intricate from a scientific point of view. He thought that the language algorithms could have detected and reproduced knowledge structures in language that had been transported there by empirical research in the first place. In other words, he thought of language as a sort of library that contains not only words, but complex statements from science. In this world view, science will enrich our vocabulary by truths as people read research and import the ideas they read into their everyday language. This is probably not possible, as language is a tool allowing us to propose and think anything and everything, and the idea is generally considered as refuted in linguistic science (Lovasz & Slaney, 2013). It actually explains why we need science to help us differ between fact, fiction and nonsense.

Still, this is exactly where there may be a way out of Smedslund’s labyrinth. The idea came to me as another colleague, on asked what he thought about the semantics project, mentioned another name that came to his mind: that of Gottlob Frege.

FREGE, WITTGENSTEIN, AND THE PROGRAMMING LANGUAGES

Gottlob Frege was a late 19th century German philosopher and logician. He is famous for three contributions to logics (Blanchette, 2012; Frege, 1884, 1918). First of all, Frege was a pioneer in creating a system of notations in formal logics that made it possible to calculate with words. Through his system, Frege was able to prove that sentences may contain degrees of similar meaning, even where the sentences do not share any words. His system was possible because he made a difference between functions in language and the arguments that the functions take. This was very important because Frege showed that there is a difference between the intrinsic logic of propositions and the content, the stuff we talk about.

The British logician George Boole had already devised a system for turning logics in to a calculating system (Boole, 1847). However, Boole's project was first and foremost a mathematical project that took the conceptual contents of propositions for given. Frege's approach was more radical. He adopted an explicit linguistic position and claimed that the meaning of a sentence resides in the proposition of the sentence, not in each single word. He wanted to create a system for calculating truths that did not stop with the logical basics, but that was also sensitive to the contents of the sentences – what the sentence is “about”, i.e., the semantic properties of propositions (Sluga, 1987).

Although his own system did not actually survive, he was an important pioneer in showing that language contains logical functions that lend themselves to complex calculations. It had originally been proposed by the 17th-century philosopher Gottfried Leibniz, who had conceived the term “*calculus ratorcinator*” (Sluga, 1987), a calculating machine that would be “...an algorithm which, when applied to the symbols of any formula of the *characteristica universalis*, would determine whether or not that formula were true as a statement of science” (Rogers, 1963, p. 934). This tradition has today evolved into programming languages, complex sets of instructions that allow computers to do efficiently and quick what was to Frege and his contemporaries long and tedious work by hand (Wiener, 1948, p. 214).

His second claim to fame came because his system was so promising that he tried to explain algebra as a branch of logics, but this effort is today judged as unsuccessful. Still, he showed that quantification and mathematical operations are strongly linked to our linguistic capabilities.

The third feature of his historical position has direct relevance to survey research. As he tried to represent the meaning of sentences through formal symbols, Frege noticed that we sometimes use different words or terms that refer to the same existing facts, but that still may convey different meanings. Consider the case of authors with pseudonyms. The three expressions “Mark Twain”, “Samuel L. Clemens” and “The author of Huckleberry Finn” all refer to the same historical person. Yet these expressions could also have slightly different meanings, one name being more tightly associated with writing while another name with a postal address or a family.

For this reason, Frege proposed a distinction between “Sinn” and “Bedeutung”, that is, meaning and reference. The three expressions above all refer to the same person, but they also have separate meanings that allow speakers to concentrate on one aspect of the person.

Frege’s logical discoveries went unheeded by the social scientists who followed Rensis Likert (1932) in exploring social realities through calculating numerical responses from surveys. A closer reflection on Frege’s claims point to the possibility that people who are apparently talking about different things, such as leadership and motivation, are really talking about the same thing, and that there will exist semantic relationships between these concepts by the way they are entered into arguments. It is these semantic relationships that create the mathematical (or statistical) relationships in the survey data. The big methodological problem was already coined at an early stage by Thorndike (1904), after whom it is called the “jingle/jangle fallacy”: In a “jingle”, there will be two groups of researchers who think they are researching different things. Closer logical scrutiny will show that they have simply developed a differing terminology, and they are actually working on the same subject. A “jangle” is the opposite, a situation where groups of researchers think they are in the same field, but their words have actually developed different references and they are no longer working on the same subject (Kelley, 1927).

A large study using semantic algorithms on the items that define constructs in social sciences was able to document the existence of widespread jingle/jangle problems in published research traditions (Larsen & Bong, 2016). The jingle/jangle fallacies are almost as predicted by Frege’s ideas, as summed up by Patricia Blanchette (2012): “...from the Fregean perspective, two sets of sentences can have radically-different syntactic properties and hence be “logically” inequivalent... while expressing exactly the same set of thoughts and hence being, from Frege’s point of view, logically indistinguishable. Similarly, two sets of sentences can be indistinguishable except for choice of atomic terms ... and yet express sets of thoughts that have, from the Fregean point of view, significantly different logical properties.”

Frege was looking for a purely propositional language that could allow a clear, unequivocal representation of a proposition or a judgement, and that would allow a comparison of how similar other expressions would be in terms of their underlying meanings.

At a time when scientists were still very much concerned with the difference between empirical and logical truths, Frege had a pupil who sought to solve this problem in a radical way. His name was Ludwig Wittgenstein and the book where he proposed his solution is called “Tractatus Logico-Philosophicus” (Wittgenstein, 1922). His main concern was to create a philosophy of science that could clarify the nature of testable empirical propositions. His main aim may not have been successful (also according to himself), but that is of no concern here. The important part for the role of semantics in survey research is that Wittgenstein and his other mentor Bertrand Russell needed to create a way to talk about language, facts and propositions.

As shown by Wittgenstein and Russell (Russell, 1922, p. 17), we can differ between different kinds of facts. Three types of facts are of particular relevance here. As a “fact”, we usually think of 1) empirical facts, as whether it is raining or not. However, the reason we want to check whether it rains or not is because we can

have different opinions on the subject. Whether someone believes it to be raining or not could be called a 2) psychological fact. But in order to believe something and discuss it, as if it is raining or not, this belief must exist in the form of a proposition that can be communicated. One may call this a 3) “logical” fact – a proposition that someone is capable of believing, or discussing with others, and ultimately check for its truth. This was central to Wittgenstein’s “mirror theory”, the assertion that there must be a systematic relationship between what we propose and the facts that we use to support or reject a theory.

Our findings when we explore survey statistics with semantics are perfectly explainable through these three types of facts. The researchers set out to explore the empirical nature of their constructs, such as “leadership” or “motivation”. They do this by obtaining records of “psychological” facts, the reported attitudes of subjects as scores on Likert scales. Eventually, when the statistics are performed, the psychological information is filtered out and the statistical patterns are no longer dependent on the individuals contributing to them. But instead of being descriptive of the empirical domain called “leadership”, the numbers are simply reflecting the semantic (or logical) relationships between the item texts.

This capability in language is the tool that helps us instruct computers today. The mechanical precursors to computers were textile producing equipment using punch cards, as shown by the engineer Charles Babbage. But as the computers got more sophisticated, they needed more systematic tools to instruct their operations, commonly referred to as “programming languages”. The pioneers of these, such as Herman Zuse, drew extensively on the groundbreaking work of logical calculations and notations developed by Frege and his British predecessor, Boole (Rojas et al., 2000; Sluga, 1987). There is an intrinsic relationship between computer languages and formal logic such that “when a [logical] specification completely defines the relations to be computed, there is no syntactic distinction between specification and program... The only difference between a complete [logical] specification and a program is one of efficiency. A program is more efficient than a specification” (Kowalski et al., 1984, p. 345). Computing languages are instructions to computers to systematically do what humans can only follow for a short while, taking full and systematic account of “what we already know”.

This is the unpleasant fact that the reviewers from the survey research tradition seem unable to realize. Our capability to detect, decode or construct logical “facts” is also tightly linked to our own meta-linguistic handicaps, and the reason why computers are useful tools that help us overcome our cognitive limitations.

One of Wittgenstein’s pressing arguments was that in order to be empirically testable, a proposition needs to be unequivocal (Russell, 1922). In Wittgenstein’s own words (1922, p. 23), “What can be said at all can be said clearly; and whereof one cannot speak thereof one must be silent.” If not, we cannot fixate the relationship between the proposition in language (the “logical fact”) and how things are (the “empirical fact”), a problem that has also been discussed by Smedslund (2002). Lack of precision in this respect creates ambiguities and discrepancies between theory and empirical observations. In other words, we must seek the strictest possible ways to fix the meaning of propositions.

Both Wittgenstein and Russell knew, and had improved on Frege's work. They were aware, not only of the computational possibilities in formal logics, but also of Frege's project trying to make the meaning of sentences primary to the logical calculus. Looming above this was also the awareness of the human limitations in making these sorts of arguments. Not only do people use language in imperfect ways, as Frege frequently pointed out but the logicians themselves become entangled in confusing conflicts that are difficult to resolve. In his foreword to the *Tractatus*, Russell (1922, p. 19) explicitly mentions that logical calculations and derivations are exceedingly difficult to follow, even for a trained mathematician: "As one with a long experience of the difficulties of logic and of the deceptiveness of theories which seem irrefutable, I find myself unable to be sure of the rightness of a theory, merely on the ground that I cannot see any point on which it is wrong." Or, as Patricia Blanchette (2012) sums up Frege's contributions: "It is hard to say what, exactly, separates a good analysis from a failed attempt." This echoes a much older lamentation from Heraclitus, the original inventor of the word "logics": That ideally, the laws of logics should be the same to everyone, even though in practice, it seems that everyone has his own (Graham, 2015).

The invention and development of logic has always followed a double-sided, almost paradoxical track: On the one hand, we are expressing ourselves in a language so precise and rule-oriented that everything we say may concomitantly invoke a host of other facts that we can infer. On the other, we easily get lost, stuck or cannot agree on these inferential steps. It is hard for us to make use of what we actually know.

Interestingly then, we have been able to create tools to help us here, precisely by turning the rules of logic into computers and programming languages. The digital algorithms are therefore giving us a possible mirror, not only to what we can achieve through logical computations, but also through exposing our lack of meta-capability. Let us turn to the text algorithms themselves.

LATENT SEMANTIC ANALYSIS AND OTHER TEXT ALGORITHMS

The close relationship between programming languages and natural languages has kept the computing community continuously interested in making computers deal with text (Schank & Abelson, 1977). Readers old enough to remember the early DOS interface of PCs also remember the cumbersome task of instructing the computer via its own language. System developers have always wanted to emulate natural languages, even after Apple and later Microsoft adopted graphical icons as substitutes for weird lines of commands.

The quest to make computers understand or produce human-like language has been labeled "Natural Language Processing" (NLP). It has made great progress in recent years as numerous digital appliances are now equipped with voice controlled interfaces. Even if the digital gadgets are not yet matching humans entirely, Apple lets you talk to its digital assistant Siri on the iPhone, a Tesla car will find addresses, call people or play music to your verbal commands, and Amazon's Alexa will talk to you about shopping. NLP is used for tasks like automatic translation, indexing of information in large bulks of texts or for easing the interface between machines and human users. Our future use of artificial intelligence (AI) will be dependent on successful NLP.

A strange obstacle for NLP has been our lack of meta-cognitive abilities as described above. The first attempts at making computers relate to natural language consisted of a chase for rules that would allow the computer to analyze or create meaning in language, such as grammar and syntax.

Some approaches to NLP still make use of such information. One such that we have been using is an algorithm termed MI (Mihalcea, Corley, & Strapparava, 2006). The MI algorithm will look up words in a lexical database called WordNet (Miller, 1995; Poli, Healy, & Kameas, 2010). WordNet is like a digital dictionary, but instead of alphabetical listings, it is a database where words are indexed for their semantic proximity to others. “Wolf” and “dog” will appear as more closely related to each other than to e.g., “ship”. In determining the meaning of a sentence, MI will identify so-called parts-of-speech and map the meanings of single words within these parts.

In this sense, MI behaves a little like a human trying to learn a foreign language – it looks up words in a dictionary (albeit an electronic one) and in a sense determines the meaning of text by taking account of their syntactic relationships.

A possibly less intuitive approach is called Latent Semantic Analysis (LSA), and was developed as a purely mathematical approach to text analysis. One of its pioneers, Thomas Landauer, even claimed that it probably simulates the way language is learnt and represented in the brain (Landauer & Dumais, 1997). While it may not be an accurate copy of the actual cerebral mechanisms, it certainly comes very close to a mathematical explanation. For this reason, some more attention will be given to LSA than to other existing algorithms. The overview of LSA given here still needs to be brief and superficial, so interested readers will have to look up the original sources to find more details (Dennis, Landauer, Kintsch, & Quesada, 2013; Gefen, Endicott, Miller, Fresneda, & Larsen, 2017).

LSA is a pure “bag-of-words”-approach, meaning that it does not use information about grammar or syntax at all. In one sense, this echoes Frege’s skepticism against using single words as sufficient containers of meaning, as he claimed that the proposition in the sentence has priority over the single words (Sluga, 1987). Instead of “knowing” meanings of words, LSA draws mathematical inferences from a huge universe of texts, called “semantic spaces”. In practice, a semantic space will have to be established by people, for example by groups of researchers. These texts may consist of thousands of excerpts from newspapers or books. The whole purpose of this text collection is to give the algorithm access to language as it is actually used by people. For example, in our own research, we have used thousands of articles from US newspapers. A semantic space is then generated from hundreds of millions of words, repeated over and over again in many contexts. The semantic space, however, is not the words themselves, but a statistical reduction applied to the relationships between all the words included in the materials.

LSA creates statistical relationships between words and the contexts in which they appear. It is this extraction of semantic relationships from the usage of words that made Landauer call LSA a mathematical theory of meaning. He thought that this process might be similar to what the brains of children do when they are exposed to use of words in the conversations of people around them (Landauer, 2007; Landauer & Dumais,

1997). LSA creates statistical relationships between words and the contexts in which they appear. In this way, the “meaning” of any word is represented as the degree to which it can replace another word in similar contexts. LSA will estimate this similarity as a number, using the following calculating steps (the reader who is uninterested in statistical analysis may skip the following paragraph):

First, LSA constructs a matrix called the “document-term”-matrix (TDM), where each row is a word and each column is a document where this word appears. This is a huge matrix in which each cell contains the number of times this word appears within each document. The TDM is then treated with a statistical technique called “Single Value Decomposition” (SVD), which is akin to factor analysis. This step turns the big matrix into three smaller ones, usually referred to as the U , Σ , and V matrices where $TDM = U \times \Sigma \times V$. These matrices contain information about the documents (U), words (V) and the singular values (Σ). The singular values are now truncated to simplify the analysis. This step is important, because the truncation determines the number of dimensions used to analyze texts later on. The result of the truncation is usually denoted as “ k ”, the number of singular values made up to describe the matrices. The number of k will determine how simplified the semantic space will be, compared to the original texts, and the significance of this will be explained further down.

LSA and similar algorithms have been used in empirical research on survey data (Arnulf et al., 2014; Arnulf, Larsen, Martinsen, et al., 2018; Arnulf, Larsen, & Martinsen, 2018; Gefen & Larsen, 2017; Nimon et al., 2015). In this case, the algorithm “projects” each item into a semantic space and estimates how it is represented in the triangular structure of $U \times \Sigma \times V$. The output is then the cosine of this relationship, a number between 0 and 1. The closer to 1, the more similar the meaning of the two terms. For the two sentences “Causes have effects” and “effects have causes”, LSA will return a cosine of 1.00 (if the reader wants to give it a try, an LSA engine can be accessed at the website lsa.colorado.edu).

In the research tradition of using Likert scales, the focus has historically been on the relationships between items or groups of items called scales. Building on the works of Cronbach and Meehl (1955), these scales have been taken as operationalizations of constructs, such as various types of “motivation”, “leadership” and similar theoretical objects. Over the years, a number of statistical procedures have been developed that analyze the quantitative properties of such scale relationships, such as Principal Component Analysis (PCA) of Structural Equation Modeling (SEM) (Jöreskog, 1993; Kline, 2005), that are purportedly able to make precise mathematical estimates of the nature of these construct relationships.

However, the Achilles’ heel of all these types of statistical modelling is that they use the co-variation between the items as their point of departure. All of them are applying correlations or covariance between the scores on the scale items as the input to the calculations. In other words, all the relationships in the models are simply iterations of the similarity among items in statistical terms.

In our research on survey statistics, we applied LSA to a series of commonly used questionnaire items. For the most part, we were able to show that the cosines computed by LSA can predict (Arnulf, Larsen, & Dysvik, 2018; Arnulf et al., 2014) and thus even replace the correlations (Arnulf, Larsen, & Martinsen, 2018). While

LSA is not as proficient as a human speaker in understanding language, it comes very close, and the “measurement scales” of the researchers have been constructed to ensure performance in the statistical models. The result of this is that the needs of the researchers and of LSA converge in the way Likert scale surveys are constructed. We have been able to re-create the PCA and SEM models using semantic information alone (Arnulf, Larsen, & Dysvik, 2018; Arnulf & Larsen, 2015; Arnulf et al., 2014), and such findings have been confirmed in independent studies (Gefen & Larsen, 2017; Nimon et al., 2015).

To put it bluntly, the statistical models of survey research will most likely reproduce the brain’s assessment of similarity between these survey items. In the language of Wittgenstein and Russell, the researchers collected information about “psychological facts” – what people believe about their bosses – to make computational models of “empirical facts” – the relationships between leadership behaviors and employee performance. Instead, they ended up with information about the “logical facts”, the numbers describing language processing in the brains of the respondents.

Almost paradoxically, the semantic algorithms provide an empirical proof of what Smedslund’s original claims (1987), as explicated in a response to a critic (1988, p. 150): “that the inter-item correlations are produced exclusively by shared logical-semantic relations, given the taken-for-granted commonsense conceptual system and the taken-for-granted contextual assumptions.”

The fundamental question is why this comes as a surprise to us, masquerading as an empirical finding that seems useful even if it only explicates what we already know. It is this incredulous resistance that keeps reoccurring in our reviewers’ rejections. It is the very same intellectual fog that Smedslund’s argumentation tries to lift.

COMPETENCE WITHOUT COMPREHENSION

But how is it possible that we know without knowing that we know?

This is a topic that has frequently been addressed in psychology as “meta-cognition”, the demonstration that we are usually much better at doing things than explaining HOW we do them. Language is itself the best case in point: While most adults are quite able speakers of their native languages, they have a much harder time explicating the rules that apply. Foreign students of German are frequently able to quote grammar rules that sound baffling to native speakers, who apply them without giving it a thought.

This phenomenon is the core point of a recent essay by the American philosopher Daniel Dennett where he compares Darwin's theory of evolution to the development of Artificial Intelligence as proposed by the logician Alan Turing (Dennett, 2012). Dennett finds that the two share a common explanation, that of “competence without comprehension”. This signifies how intelligent systems develop capabilities that the system itself cannot explain. In fact, from a computational point of view, the output of the computations usually show no resemblance to the machinery that brought the computations about.

Specifically, the DNA code of species can be compared to computer algorithms. Alan Turing lay the foundations of computing science in 1936 by proving that “It is possible to invent a single machine which can

be used to compute any computable sequence.” The building blocks of the Turing machine were simple pieces of information (0’s and 1’s) with rules of combinations, very much inspired by the works of Frege (Beeson, 2004, p. 6). In the same way, the DNA molecule stores and expresses information by long combinations of the simple base-pairs of G-C (guanine-cytosine) and A-T (adenine-thymine).

In other words, observable biological phenomena – such as the brain’s ability to produce language – are products of calculations, but the calculations themselves are usually not apparent to the speakers.

The experience of invariant calculations still appears to the speakers now and then. The notion of “logic” is one such phenomenon. The Greek philosopher Heraclitus living around 500 BC is usually credited with coining the term. He observed how the universe seemed structured as a universally consistent language, because there seems to be a lawful consistency in meticulous descriptions of nature. As he pointed out, the way up and the way down is the same way. It was our tendency to lose this out of sight (and hence the need to remind us of their identity) that made him issue the warning already quoted above, that although “this Word is common, the many live as if they had a private understanding” (Graham, 2015).

This seemingly dual nature of logic has haunted our intellectual efforts ever since: On the one hand, there appears to be an independent lawfulness of the relationships of words and expressions to each other. On the other hand, it is as if the individual always struggles and frequently fails to live by these rules. Although as children we are quick to absorb and use the regularities of language, most of us struggle to use them perfectly. And, most importantly, we seem not to entirely grasp the full implication of the logical linkages that language provides, as per Russell’s comment in his foreword to Wittgenstein’s *Tractatus*, that a stringent scrutiny of a logical theorem was tough even to a trained logician.

This struggle has kept philosophy in a continuous pendulum between logical rationalism – the claim that observation is unnecessary as most problems can be solved through thinking – to theory-rejecting empiricism that distrusts products of the mental apparatus, trusting only what can be measured (Markie, 2017). One core proposition in Smedslund’s work is that psychology will always be entangled in the intricacies between logical and empirical questions, where researchers keep looking for empirical questions, only to re-discover what was logically necessary.

This is where I believe that our discoveries using text algorithms may help us forward. Text algorithms like LSA take a purely calculative approach. Even if these calculations themselves take only seconds in a prepared semantic space, they may model the way a child’s brain calculates the meanings of words during the years of exposure to its native language. Landauer already pointed out how LSA can solve “Plato’s paradox” – the fact that children can know so many words for things that they have never actually encountered in real life (Landauer & Dumais, 1997). These words are calculable from their semantic networks with other words. An increasing vocabulary implies an increasing differentiation and resolution of details.

In the tradition of Frege and Wittgenstein, it is interesting to ask the seemingly hopeless question: “How many things are there in the universe?” The answer is that it depends on the respondent’s conceptual richness. A simple answer may be that there is only one – the universe. Any attempt at specifying more

numbers will depend on words that differentiate – round things, blue things, heavy things, small things and so on.

The practical implication of this is that our level of details in linguistic competence may drown speakers in the details of language, losing its inherent calculative relationships out of sight. Because, as my son once pointed out to me, “there may be many things in the universe that do not have words attached to them, but all words will also be related to other words.” To be a meaningful word, any word needs to be defined in terms of others. Our language is thus a huge semantic network where all words are by necessity logically linked to others, however distantly. As our vocabulary increases, we can keep reiterating statements and fall victim to the idiosyncrasies as noted by Heraclitus and Russell and finally look bewilderedly for empirical facts to support our arguments and settle our disputes. We are locked inside Smedslund’s labyrinth.

One may think of our semantic network as an enormous crossword puzzle where all words are fixed in their mutual relationships. With our cognitive constraints, we cannot see this – which is why most people find crosswords difficult to solve when the fields are empty, but recognizable as correct when the letters are filled in. In reality, it may be more like a giant Sudoku, where the meaning of any expression will be mathematically fixed by its relationship to all other measures. Psychological theories, then, are frequently not theoretical generalizations of empirical observations. Instead, they may simply be logical iterations of already given propositions. As a theory is argued by its authors, the concepts involved are defined in terms of each other, and the relationships become self-evident or tautological (Semin, 1989; Smedslund, 1988, 1994, 2015; van Knippenberg & Sitkin, 2013). The authors and their readers are unaware of the fact that they are merely iterating truths given by the conditions. Like solvers of crossword puzzles, they do not see the solution as self-evident, but simply sense their own cognitive effort paired with a feeling that the line of thinking is reasonable.

At this point I want to return to the issue of the k dimensions in LSA, as described in the section above. If the number of k is very low, the LSA algorithm will tend to simplify everything and estimate higher degrees of similarity between texts, such as sentences. If the number of k is very high, the algorithm may fail to detect similarities until texts become very similar.

Consider the following examples:

If we enter the sentences “Your dog is loose and runs around”, “Your hound is roaming about”, and “A rabbit sleeps in its hole”, the LSA algorithm will detect the differences between them. If we set k to 300, the algorithm will find the sentences with the synonyms “hound” and “dog” very similar, as their cosine will be .75, while only .40 or .33 with the sentence about a sleeping rabbit. However, if we reduce the number of k to 10, the similarity between the two first sentences increases to .95, but the rabbit is now also estimated at .82 with the sentence about the dog. It is as if LSA looks meticulously at sentences and determines that they are related but not the same when k is set to 300. When k is reduced to 10, LSA seems to make less differentiated, almost sloppy judgments – these are all sentences about some kind of animal in a location.

<i>Using k=300</i>	Your dog is loose and runs around	Your hound is roaming about	A rabbit sleeps in its hole
Your dog is loose and runs around	1	0.75	0.40
Your hound is roaming about	0.75	1	0.33
A rabbit sleeps in its hole	0.40	0.33	1

<i>Using k=10</i>	Your dog is loose and runs around	Your hound is roaming about	A rabbit sleeps in its hole
Your dog is loose and runs around	1	0.95	0.82
Your hound is roaming about	0.95	1	0.67
A rabbit sleeps in its hole	0.82	0.67	1

The effects of the differences k dimensions of LSA are reminiscent of the jingle/jangle fallacies mentioned above, where similar concepts exist under different names, and similar names refer to very different concepts. It is also relevant to Frege’s distinction between Sinn and Bedeutung (meaning and reference): The precise meaning of a word in the sense of its reference may in practice be a matter of precision. A roaming hound may mean something different from a running dog. Depending on the context, it may also mean the same – even being similar to a rabbit sleeping in a hole.

This calculative capacity of language is exercised whenever we are trying to solve a crossword puzzle. Expressions may mean the same or be distinct, but it frequently requires an intellectual effort to determine this as the calculations of linguistics do not always come as effortless options (Kahneman, 2011).

The semantic calculations of the brain are remarkably flexible and precise at the same time. It seems that they are capable of loosening the semantics restrictions almost entirely, as when forming poetry and allegories. The meaning of an allegory is precisely not what it is “about”, as in Shakespeare’s famous sonnet: “Shall I compare thee to a summer’s day?” We can enter this in LSA (helping the American algorithm by replacing “thee” with “you”), and test its similarity with two interpretations: One is a poetic transcription, “I find you warm, bright, and lovely”, the other a more concrete explication: “Your name may be June”. Although LSA sees a possibility that Shakespeare is addressing someone named June (cosine = .40), it finds it more likely that the poet refers to the personality of the interlocutor (cosine = .67).

Our linguistic capabilities are thus at the same time a product of precise and complex calculations but also leave us mostly aware of probabilistic results with wide room for error and individual interpretations. Being competent without comprehension, in Dennett’s words, we find ourselves locked in a labyrinth of semantic networks that appear as logical lawfulness, without being able to overlook it.

Our languages are collective, cultural accumulations of words in which all statements need to be implicitly locked into all other statements to be intelligible. The individual does not have access to this complexity due

to lack of cognitive capacity – we merely have competence, but not comprehension. In the statistical models created by the responses to Likert-scale items, the machinery of the competence re-appears as patterns of correlations. This is an instance of “the wisdom of crowds”, because it will be the mean response pattern that carries the signal. Individual response protocols seem to contain a lot of semantic noise, as Heraclitus would have recognized.

In our data, it usually takes a few hundred respondents to approximate the structures suggested by the algorithms. If we use only native speakers of English, they will approximate the LSA results quicker than speakers of other languages, but hundreds of Norwegians and even Chinese eventually arrive at the same quantitative structures as predicted by algorithms in American English.

There may even be a linguistic relativity phenomenon in here somewhere: Chinese responding in Chinese are slower to approach the LSA-predicted semantics than Chinese in English. Chinese as a language is far looser in its semantic restrictions than indo-european languages (Harbsmeier, 2007), while Germans responding in German seem to comply with the LSA-predicted semantics far quicker than even native English speakers. That may be one reason why German speakers like Frege and Wittgenstein were pioneers in analytical philosophy, and why Chinese does not even actually have an indigenous word for “logic” (Nisbett, Peng, Choi, & Norenzayan, 2001; Norenzayan, Smith, Kim, & Nisbett, 2002). Instead, ancient Chinese philosophy articulated a skepticism towards language as a tool, seeing that it has only limited capability to contain truths about the world (Feng, 2015). Some languages may simply structure the output in ways that makes the computational underpinnings more obvious to the speaker than others, making the ancient Greeks like Plato embrace idealism while the Chinese discarded it.

WITTGENSTEIN’S REVENGE AS A WAY OUT OF SMEDSLUND’S LABYRINTH

I have titled this chapter “Wittgenstein’s revenge” because despite his and Russell’s fame in the 1920’s, their call for more stringent philosophical cleaning of research questions went unheeded, at least in psychology. While the behaviorist movement did call for a more skeptical treatment on non-observable phenomena, these were re-introduced from physics (Bridgman, 1927) through the concept of “operationalism” (Boring, 1945). Operationalism allowed constructs to be defined through the procedures used to measure them. This instigated Cronbach and Meehl to introduce a 50 year long hegemony of empiricism, sanctioned explicitly by the methodological conventions of the American Psychological Association (AERA, APA, & NCME, 2014; APA, 1954; Kathleen L. Slaney, 2017; K. L. Slaney & Racine, 2013).

This empiricism gained momentum from the increasing access to advanced statistical models in computing that made factor analysis and structural equations the preferred tools of any researcher who wanted to gain tenure in quantitative research. The need to resort to painful philosophical reflections on the empirical versus logical nature of the research questions seemed to be omitted. One could simply turn any question into a 7-point Likert scale, gather responses and begin the computing. It did not, and still does not seem to matter that the nature of the numbers – the *what* of what’s being measured – is usually not a part of the discussion and

harder to explain than the statistical operations themselves (Lamiell, 2013; Mari, Maul, Irribarra, & Wilson, 2017; Maul, 2017).

It is therefore ironic that the main heritage of Boole, Frege, Wittgenstein and their contemporaries was kept alive in the computing tools themselves – in hardware as well as in the software. As all human work processes are increasingly becoming subject to digitalization, the original projects of the logician pioneers seem reintroduced into the research process itself. The phrase “Wittgenstein’s revenge” may be overly catchy, but I believe there is an opportunity to re-appraise his tradition in empirical research through the digitized tools of formal logics (hence the idea that he is coming back with a vengeance).

At first glance, it may seem as if our empirical research, in supporting Smedslund’s argumentation, may be just as much a vindication of Frege. However, I think there is a line of development from Boole through Frege to Wittgenstein that is so far unexploited in psychology. Boole saw that logical operations could be formalized into computations. From there, Frege moved on from mere operators to the calculated analysis of propositions – analyzing not only logical, but semantic relationships. Finally, while he recognized these previous attempts, Wittgenstein was not satisfied with remaining in the field of logic. He raised the question about the limits of language as a container of scientific knowledge, saying that “In logic nothing is accidental: if a thing can occur in an atomic fact the possibility of that atomic fact must already be prejudged in the thing.” (Wittgenstein, 1922, prop. 2.012). Further: “The proposition is not a mixture of words (just as the musical theme is not a mixture of tones)” (prop. 3.141). Words cannot be haphazardly blended, but will only be meaningfully combined in combining the logical/semantic properties that are already given in the definitions of the words themselves. The possible combinations of relationships are vast, but in themselves fixed. Wittgenstein located the “mysterious” in realities that certainly exist but that defy logical description, and famously warned against discussing it. This is a locked universe of meaning that we cannot escape.

Or maybe we can. Russell commented (1922, p. 18) that “after all, Mr Wittgenstein manages to say a good deal about what cannot be said, thus suggesting to the skeptical reader that possibly there may be some loophole through a hierarchy of languages, or by some other exit.” One reason for our lack of escape from Smedslund’s labyrinth has probably been our lack of an impartial, third party judgement of logical or semantic truths. Now that the algorithms have come closer than ever to Leibniz’s dream of the “*calculus ratorator*”, they could provide a tool for exploring the no-man’s land between the semantic and the empirical, targeting and describing our cognitive barriers.

Tool-making has helped humans overcome many types of shortcomings before, increasing our physical strength and our traveling capabilities. As we are improving our cognitive tools, we may also be expanding our empirical reach into what was earlier the exclusive realm of philosophy. As we improve our capability to apply digital analytics not only to the observations, but to our theories and research questions themselves, we may be making real progress in differing between logical and empirical questions.

It may also help us explore the fascinating details of why we fail to comply with semantic and logical guidelines. A growing body of psychological knowledge has documented our cognitive shortcomings and

driven the notion of “rational man” out of economics, a field covering two Nobel prizes in economics (Kahneman, 2011; Simon, 1957; Todd & Gigerenzer, 2003; B. Tversky & Kessell, 2014). The semantically expected is not uninteresting, whether in itself as documenting the brain’s seemingly effortless and yet very precise linguistic parsing capabilities (Michell, 1994), or even more as an impartial yardstick for assessment of our failure to comply (Gebotys & Claxton-Oldfield, 1989; Kahneman & Tversky, 1973; A. Tversky & Kahneman, 1974).

CONCLUSION: DOES IT MATTER?

This chapter started out describing the disbelief of reviewers confronted with the fact that their research objects were predictable a priori. My interpretation of their individual reactions was that they were being “competent without comprehension”. The bigger challenge – that of the scientific community – has been its entanglement in a failure to recognize the difference between logical and empirical problems as described by Smedslund. As we and other researchers have shown repeatedly in the recent years, we now actually may have the tools that could help us explore these questions, clear unnecessary confusion, and make way for real progress in psychology.

As a small practical example towards the end, I just want to share the way that I personally apply this new type of knowledge as a practical approach to one of my teaching fields, leadership development.

During the introductory part of session on leadership with experienced managers, I will frequently introduce myself as a researcher on leadership. I then ask the audience if they think it meaningful to do research on whether good leadership creates better results in organizations. The usual response is a solemn accept of this kind of research. I ask them to define “leadership”, and most definitions they come up with contain “results” in them, typically in the form of “achieving goals by co-operation” or something like it. In that case I say, they should also endorse doing research on what it is about Mondays that creates Tuesdays. If “achievement” is part of the definition, one cannot research whether leadership creates some kind of achievement. We have already decided that as part of the definition (van Knippenberg & Sitkin, 2013).

One could easily ridicule the management field for falling victim to thoughtless fads and types of “consultant speech”, but this fails to recognize the more important point that we are all competent without comprehension. We become trapped in real problems and get locked inside versions of Smedslund’s labyrinth by being competent without comprehension. The resistance of the reviewers when faced with these possibilities may have been fueled with a sense of rejection, that the efforts were all in vain as instances of “hard obscurantism” (Elster, 2011).

I believe that the human mind is locked in behind its own cognitive limitations. These limitations may not have played a big role in the natural habitat where homo sapiens emerged. As we have placed ourselves in an increasingly complex system of behavioral, technological and economic feedback loops, there may be a real need for us to understand these limitations (Harari, 2015; Senge, 2000; Soros, 2006). Our digital crutches are

evolving fast and playing into most areas of social decision-making. Psychological research aimed at understanding how our cognitive limitations relate to our new tools will hopefully contribute to keep the developing technology a servant instead of a master.

REFERENCES

- AERA, A. E. R. A., APA, A. P. A., & NCME, N. C. o. M. i. E. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- APA, A. P. A. (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, *51*, 1-38.
- Arnulf, J. K., Larsen, K., & Dysvik, A. (2018). Measuring Semantic Components in Training and Motivation: A Methodological Introduction to the Semantic Theory of Survey Response. *Human Resource Development Quarterly*, *Forthcoming*.
- Arnulf, J. K., & Larsen, K. R. (2015). Overlapping semantics of leadership and heroism: Expectations of omnipotence, identification with ideal leaders and disappointment in real managers. *Scandinavian Psychologist*, *2*(e3). doi:10.15714/scandpsychol.2.e3
- Arnulf, J. K., Larsen, K. R., Martinsen, O. L., & Bong, C. H. (2014). Predicting survey responses: how and why semantics shape survey statistics on organizational behaviour. *PLoS ONE*, *9*(9), e106361. doi:10.1371/journal.pone.0106361
- Arnulf, J. K., Larsen, K. R., Martinsen, O. L., & Egeland, T. (2018). The failing measurement of attitudes: How semantic determinants of individual survey responses come to replace measures of attitude strength. *Behav Res Methods*, 1-21. doi:10.3758/s13428-017-0999-y
- Arnulf, J. K., Larsen, K. R., & Martinsen, Ø. L. (2018). Respondent Robotics: Simulating Responses to Likert-Scale Survey Items. *Sage Open*, *8*(1), 1-18. doi:10.1177/2158244018764803
- Avolio, B. J., Bass, B. M., & Jung, D. I. (1995). *Multifactor Leadership Questionnaire technical report*. Redwood City, CA.: Mind Garden.
- Bagozzi, R. P. (2011). Measurement and Meaning in Information Systems and Organizational Research: Methodological and Philosophical Foundations. *MIS Quarterly*, *35*(2), 261-292. doi:<https://doi.org/10.2307/23044044>
- Beeson, M. J. (2004). The Mechanization of Mathematics. In C. Teuscher (Ed.), *Alan Turing: Life and Legacy of a Great Thinker* (pp. 1-54). Berlin: Springer Verlag.
- Blanchette, P. A. (2012). *Frege's conception of logic*
- Boole, G. (1847). *The Mathematical Analysis of Logic, Being an Essay towards a Calculus of Deductive Reasoning*. London, England: Macmillan, Barclay, & Macmillan.
- Boring, E. G. (1945). The Use of Operational Definitions in Science. *Psychological Review*, *52*(5), 243-245. doi:DOI 10.1037/h0054934
- Bridgman, P. W. (1927). *The Logic of Modern Physics*. New York: Macmillan.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychol Bull*, *52*(4), 281-302. doi:<https://doi.org/10.1037/h0040957>
- Dennett, D. (2012). A Perfect and Beautiful Machine': What Darwin's Theory of Evolution Reveals About Artificial Intelligence. *The Atlantic*.
- Dennis, S., Landauer, T. K., Kintsch, W., & Quesada, J. (2013). *Introduction to latent semantic analysis*. University of Colorado.
- Elster, J. (2011). Hard and Soft Obscurantism in the Humanities and Social Sciences. *Diogenes*, *58*(1-2), 159-+. doi:10.1177/0392192112444984
- Feng, Y. (2015). *A short history of Chinese philosophy*. Beijing: Foreign language teaching and research press.
- Frege, G. (1884). *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Breslau: W. Koebner.
- Frege, G. (1918). Der Gedanke. Eine logische Untersuchung *Beiträge zur Philosophie des deutschen Idealismus I* (pp. 58-77).

- Gebotys, R. J., & Claxton-Oldfield, S. P. (1989). Errors in the quantification of uncertainty - a product of heuristics or minimal probability knowledge base. *Applied Cognitive Psychology*, 3(3), 237-250.
- Gefen, D., Endicott, J. E., Miller, J., Fresneda, J. E., & Larsen, K. R. (2017). A Guide to Text Analysis with Latent Semantic Analysis in R with Annotated Code: Studying Online Reviews and the Stack Exchange Community. *Communications of the Association for Information Systems*, 41(11), 450-496.
- Gefen, D., & Larsen, K. (2017). Controlling for Lexical Closeness in Survey Research: A Demonstration on the Technology Acceptance Model. *Journal of the Association for Information Systems*, 18(10), 727-757. doi:<https://doi.org/10.17705/1jais.00469>
- Graham, D. W. (2015). Heraclitus. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition ed.).
- Harari, Y. N. (2015). *Sapiens: A Brief History of Humankind*. New York: Harper Collins.
- Harbsmeier, C. (2007). *Globalisation and conceptual biodiversity*. Paper presented at the Union Académique Internationale,, Oslo.
- Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 294-316). Newbury Park: Sage.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237-251.
- Kelley, T. L. (1927). *Interpretation of educational measurements*. New York, NY: World Book.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2 ed.). New York: Guilford Press.
- Kowalski, R., Jackson, M. A., Rogers, M. J., Shepherdson, J. C., Sannella, D., & Lehman, M. M. (1984). The Relation between Logic Programming and Logic Specification [and Discussion]. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 312(1522), 345-361. doi:10.1098/rsta.1984.0064
- Kuvaas, B. (2006). Performance appraisal satisfaction and employee outcomes: mediating and moderating roles of work motivation. *International Journal of Human Resource Management*, 17(3), 504-522. doi:10.1080/09585190500521581
- Lamiell, J. T. (2013). Statisticism in personality psychologists' use of trait constructs: What is it? How was it contracted? Is there a cure? *New Ideas in Psychology*, 31(1), 65-71. doi:10.1016/j.newideapsych.2011.02.009
- Landauer, T. K. (2007). LSA as a theory of meaning. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 3-34). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211-240. doi:Doi 10.1037//0033-295x.104.2.211
- Larsen, K. R., & Bong, C. H. (2016). A Tool for Addressing Construct Identity in Literature Reviews and Meta-Analyses. *MIS Quarterly*, 40(3), 529-+. doi:Doi 10.25300/Misq/2016/40.3.01
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*, 140, 1-55.
- Lovasz, N., & Slaney, K. L. (2013). What makes a hypothetical construct "hypothetical"? Tracing the origins and uses of the 'hypothetical construct' concept in psychological science. *New Ideas in Psychology*, 31(1), 22-31. doi:10.1016/j.newideapsych.2011.02.005
- MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct Measurement and Validation Procedures in Mis and Behavioral Research: Integrating New and Existing Techniques. *MIS Quarterly*, 35(2), 293-334.
- March, J. G., & Simon, H. A. (1958). *Organizations*. New York: Wiley.

- Mari, L., Maul, A., Irribarra, D. T., & Wilson, M. (2017). Quantities, Quantification, and the Necessary and Sufficient Conditions for Measurement. *Measurement, 100*, 115-121. doi:10.1016/j.measurement.2016.12.050
- Markie, P. (2017). Rationalism vs. Empiricism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017 ed.). Stanford, CA: Stanford University.
- Maul, A. (2017). Rethinking Traditional Methods of Survey Validation. *Measurement-Interdisciplinary Research and Perspectives, 15*(2), 51-69. doi:10.1080/15366367.2017.1348108
- Michell, J. (1994). Measuring Dimensions of Belief by Unidimensional Unfolding. *Journal of mathematical psychology, 38*(2), 244-273. doi:DOI 10.1006/jmps.1994.1016
- Michell, J. (2013). Constructs, inferences, and mental measurement. *New Ideas in Psychology, 31*(1), 13-21. doi:10.1016/j.newideapsych.2011.02.004
- Mihalcea, R., Corley, C., & Strapparava, C. (2006). Corpus-based and knowledge-based measures of text semantic similarity. *AAAI, 6*, 775-780.
- Miller, G. A. (1995). Wordnet - a Lexical Database for English. *Communications of the ACM, 38*(11), 39-41. doi:Doi 10.1145/219717.219748
- Nimon, K., Shuck, B., & Zigarmi, D. (2015). Construct Overlap Between Employee Engagement and Job Satisfaction: A Function of Semantic Equivalence? *Journal of Happiness Studies, 17*(3), 1149-1171. doi:10.1007/s10902-015-9636-6
- Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: holistic versus analytic cognition. *Psychol Rev, 108*(2), 291-310.
- Norenzayan, A., Smith, E. E., Kim, B. J., & Nisbett, R. E. (2002). Cultural preferences for formal versus intuitive reasoning. *Cognitive Science, 26*(5), 653-684. doi:10.1207/s15516709cog2605_4
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. In S. T. Fiske, D. L. Schacter, & S. E. Taylor (Eds.), *Annual Review of Psychology, Vol 63* (Vol. 63, pp. 539-569). Palo Alto: Annual Reviews.
- Poli, R., Healy, M., & Kameas, A. (2010). WordNet. In C. Fellbaum (Ed.), *Theory and applications of ontology: Computer applications* (pp. 231-243). New York: Springer.
- Rogers, H. (1963). An Example in Mathematical Logic. *American Mathematical Monthly, 70*(9), 929-&. doi:Doi 10.2307/2313050
- Rojas, R., Göktekin, C., Friedland, G., Krüger, M., Langmack, O., & Kuniss, D. (2000). *Plankalkül: The First High-Level Programming Language and its Implementation*. Retrieved from Berlin:
- Russell, B. (1922). An introduction to the tractatus logico-philosophicus. In L. Wittgenstein (Ed.), *Tractatus logico-philosophicus*. London: Kegan Paul.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. New York: Routledge.
- Semin, G. (1989). The contribution of linguistic factors to attribute inference and semantic similarity judgements. *European Journal of Social Psychology, 19*, 85-100.
- Senge, P. (Producer). (2000). The Leadership of Profound Change. *SPC INK*. Retrieved from http://66.102.9.104/search?q=cache:elAoLNjDcgJ:www.spcpress.com/ink_pdfs/Senge.pdf+SPC+INK,+2000,+%231.&hl=en&ct=clnk&cd=1&gl=no
- Simon, H. A. (1957). *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: Wiley.
- Slaney, K. L. (2017). *Validating Psychological Constructs : Historical, Philosophical, and Practical Dimensions*. London, UK: Palgrave MacMillan.
- Slaney, K. L., & Racine, T. P. (2013). Constructing an understanding of constructs. *New Ideas in Psychology, 31*(1), 1-3. doi:10.1016/j.newideapsych.2011.02.010
- Sluga, H. (1987). Frege Against the Booleans. *Notre Dame Journal of Formal Logic, 28*(1).

- Smedslund, J. (1978). Banduras Theory of Self-Efficacy - Set of Common-Sense Theorems. *Scandinavian Journal of Psychology*, 19(1), 1-14. doi:DOI 10.1111/j.1467-9450.1978.tb00299.x
- Smedslund, J. (1987). The epistemic status of inter-item correlations in Eysenck's Personality Questionnaire: the a priori versus the empirical in psychological data. *Scandinavian Journal of Psychology*, 28, 42-55. doi:<https://doi.org/10.1111/j.1467-9450.1987.tb00904.x>
- Smedslund, J. (1988). What Is Measured by a Psychological Measure. *Scandinavian Journal of Psychology*, 29(3-4), 148-151. doi:DOI 10.1111/j.1467-9450.1988.tb00785.x
- Smedslund, J. (1994). Nonempirical and Empirical Components in the Hypotheses of 5 Social-Psychological Experiments. *Scandinavian Journal of Psychology*, 35(1), 1-15. doi:DOI 10.1111/j.1467-9450.1994.tb00928.x
- Smedslund, J. (1995). Psychologic: Commonsense and the pseudoempirical. In J. Smith, R. Harre, & L. Van Langenhove (Eds.), *Rethinking Psychology* (pp. 196-206). London: Sage.
- Smedslund, J. (2002). From hypothesis-testing psychology to procedure-testing psychologic. *Review of General Psychology*, 6, 51-72.
- Smedslund, J. (2012). Psycho-logic: Some thoughts and after-thoughts. *Scandinavian Journal of Psychology*, 55, 295-302.
- Smedslund, J. (2015). The Value of Experiments in Psychology *The Wiley Handbook of Theoretical and Philosophical Psychology* (pp. 359-373). New Jersey: John Wiley & Sons, Ltd.
- Smedslund, J. (2016). Why Psychology Cannot be an Empirical Science. *Integr Psychol Behav Sci*, 50(2), 185-195. doi:10.1007/s12124-015-9339-x
- Soros, G. (2006). *The age of fallibility: The consequences of the war on terror*. London: Weidenfeld & Nicolson.
- Stogdill, R. M. (1963). *Manual for the Leader Behavior Description Questionnaire, form XII*. Columbus, OH: Bureau of Business Research, Ohio State University.
- Thorndike, E. (1904). *An introduction to the theory of mental and social measurements*. New York: Columbia university: Teachers college.
- Todd, P. M., & Gigerenzer, G. (2003). Bounding rationality to the world. *Journal of Economic Psychology*, 24(2), 143-165. doi:10.1016/S0167-4870(02)00200-3
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-1131. doi:10.1126/science.185.4157.1124
- Tversky, B., & Kessell, A. (2014). Thinking in action. *Pragmatics & Cognition*, 22(2), 206-223. doi:10.1075/pc.22.2.03tve
- van Knippenberg, D., & Sitkin, S. B. (2013). A Critical Assessment of Charismatic—Transformational Leadership Research: Back to the Drawing Board? *The Academy of Management Annals*, 7(1), 1-60. doi:10.1080/19416520.2013.759433
- van Schuur, W. H., & Kiers, H. A. L. (1994). Why factor analysis often is the incorrect model for analyzing bipolar concepts, and what models to use instead. *Applied psychological measurement*, 18(2), 97-110.
- Wiener, N. (1948). *Cybernetics*. New York: Wiley.
- Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. London: Kegan Paul.
- Yukl, G. (2012). *Leadership in organizations (8th ed.)*. Harlow, UK: Pearson Education.