Year : 2022

Three Essays in Management: Methods, Empirics,History and Critique

Kleinbauer Tyler

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE COMPORTEMENT ORGANISATIONNEL

Three Essays in Management: Methods, Empirics, History and Critique

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences Économiques, mention « Management »

par

Tyler KLEINBAUER

Directeur de thèse
Prof. John Antonakis

Jury

Prof. Rafael Lalive, Président
Prof. Christian Zehnder, expert interne
Prof. Jeffrey Edwards, expert externe
Prof. Mikko Rönkkö, expert externe

LAUSANNE
2022

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE COMPORTEMENT ORGANISATIONNEL

Three Essays in Management: Methods, Empirics, History and Critique

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences Économiques, mention « Management »

par

Tyler KLEINBAUER

Directeur de thèse
Prof. John Antonakis

Jury

Prof. Rafael Lalive, Président
Prof. Christian Zehnder, expert interne
Prof. Jeffrey Edwards, expert externe
Prof. Mikko Rönkkö, expert externe

LAUSANNE
2022

UNIL | Université de Lausanne
HEC Lausanne
Le Décanat
Bâtiment Internef
CH-1015 Lausanne

IMPRIMATUR

Sans se prononcer sur les opinions de l'auteur, la Faculté des Hautes Etudes Commerciales de l'Université de Lausanne autorise l'impression de la thèse de Monsieur Tyler Richard KLEINBAUER, titulaire d'un baccalauréat universitaire ès Sciences en Management de l'Université de Lausanne, titulaire d'une maîtrise universitaire ès Sciences en Management de l'Université de Lausanne, en vue de l'obtention du grade de docteur ès Sciences économiques, mention « management ».

La thèse est intitulée :

THREE ESSAYS IN MANAGEMENT: METHODS, EMPIRICS, HISTORY AND CRITIQUE

Lausanne, le 29 septembre 2022

La Doyenne

Marianne SCHMID MAST

HEC Lausanne
Le Décanat
Tél. +41 21 692 33 40 | Fax +41 21 692 33 05
www.hec.unil.ch | hecdoyen@unil.ch

Members of the Thesis Committee

Prof. John Antonakis

Professor of Organizational Behavior – University of Lausanne, Switzerland

Thesis supervisor

Prof. Christian Zehnder,

Professor of Organizational Decision Making – University of Lausanne, Switzerland

Internal Expert of the thesis committee

Prof. Mikko Rönkkö

Associate Professor of Entrepreneurship – Jyväskylä University School of Business and Economics, Finland

External Expert of the thesis committee

Prof. Jeffrey Edwards,

Belk Distinguished Professor of Organizational Behavior – UNC Kenan-Flagler, USA

External Expert of the thesis committee

University of Lausanne
Faculty of Business and Economics

Ph.D. in Economics
Subject area "Management"

I hereby certify that I have examined the doctoral thesis of

Tyler KLEINBAUER

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ______________________ Date: 04/09/2022

Prof. John ANTONAKIS
Thesis supervisor

University of Lausanne
Faculty of Business and Economics

Ph.D. in Economics
Subject area "Management"

I hereby certify that I have examined the doctoral thesis of

Tyler KLEINBAUER

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ______________________________ Date: 05.09.2022

Prof. Christian ZEHNDER
Internal member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

Ph.D. in Economics
Subject area "Management"

I hereby certify that I have examined the doctoral thesis of

Tyler KLEINBAUER

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ______________________ Date: 16/9/2022

Prof. Mikko RÖNKKÖ
External member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

Ph.D. in Economics
Subject area "Management"

I hereby certify that I have examined the doctoral thesis of

Tyler KLEINBAUER

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: Jeff Edwards Date: 13 September 2022

Prof. Jeffrey EDWARDS
External member of the doctoral committee

Three Essays in Management: Methods, Empirics, History and Critique

Tyler R. Kleinbauer

Acknowledgements

I want to thank my supervisor, John Antonakis, for his trust, feedback and mentorship throughout my PhD. John has always been available for guidance, and this despite his many other responsibilities. John has taught me how to identify, sharpen and realize research opportunities. He is also one of the best communicators I've had the pleasure of working with. Orally, he is charismatic, enthusiastic and entertaining, whereas in writing he is sharp, persuading, and has a sound vision of where the text should go. He also gave me the space to develop myself as an independent thinker. Working with John has been a fantastic experience and I've learned much from him. Overall, the skills I've acquired are transversal, applicable in any domain, and will serve me throughout my life.

I also want to thank Mikko Rönkkö for teaching me the craft of clear, precise and concise writing, through many rounds of pointed feedback on my first chapter. Because writing is in some sense the highest form of thinking, Mikko has helped me become a better thinker, which is something I value highly. Mikko has an attention to detail that I can only ever aspire to attain and has helped me bring my thesis to a level I would have not been able to attain alone. I also thank Jeffrey Edwards and Christian Zehnder for their careful read of my thesis, and pointed comments which helped improve my thesis.

I want to thank my wife, Leïla, for proofreading, brainstorming and listening to me talk about my thesis: she must know my thesis better than I at this point. She's been incredibly supportive throughout these years and has helped me maintain my sanity through the ups and downs of accomplishing a PhD.

I also want to thank my friends and colleagues, in particular, Anely, Kostas, Mahshid, Mathieu, Matthieu, Moritz, Morten, Nicolas, Sirio, Thomas, and Tristan for the fun times, interesting conversations and advice throughout the years.

Last but not least, I want to thank my family and friends, for their support.

Synthesis Report

In my first chapter, we investigate the use and utility of a family of statistical techniques called relative importance analysis. Relative importance analysis techniques are increasingly used in organizational research to assess the importance of various causes on their outcomes. These techniques are attractive because they promise to assess the importance of variables unambiguously and objectively. Regardless of their popularity, these techniques and their use have not been thoroughly scrutinized. We review the literature to see how relative importance techniques are used and critically assess if these tools are effective in addressing the problems they are used for. We find that researchers predominantly use relative importance techniques to assess the relative causal importance of variables, which these techniques are unsuitable for. We back our arguments with intuitive explanations, formal analysis of the relative importance techniques, and a Monte Carlo simulation comparing relative importance techniques against regression analysis. We show that regression provides a more accurate ranking of the relative causal importance of variables across a variety of scenarios. We conclude with practical guidelines on how to assess the relative causal importance of variables using the straightforward idea of comparable investments.

In the second chapter of my thesis, we seek to settle the decades old debate about the social construction of leadership. Societies are facing numerous grand challenges and leaders are increasingly counted on to provide solutions. But can top-level leaders affect outcomes that unfold via various pathways? Some research streams have suggested that organizational outcomes may not be caused by, but are simply ascribed to the leader; leadership may merely be a consequence of causal attributions. We provide a rigorous test to determine whether leaders matter by exploiting a very controlled, though unusual leadership context, where leader discretion is restricted; that of U.S. state governorships. This context allows us to estimate precisely what role top-level leaders and their teams may play in determining institutional outcomes, measured on a standard metric. We quantify the "leadership effect" in a sample of 500 governors across the 50 states of the U.S. and the district of Columbia. We implement state-of-the-science methodical advances in variance decomposition on a sample of 2,985 governorship-time observations, covering the periods 1963 to 2019, to explain variance in real yearly Gross Domestic Product (GDP) growth. After having partialed out time effects (0.47 of the variance in real yearly GDP growth), we show that governors and their administrations are responsible for 2.36% of the variation in real GDP growth, while state-effects only account for 0.77% of the variation in real yearly GDP growth. Our results contradict earlier research suggesting that top-level leadership may not matter (e.g., Salancik & Pfeffer, 1977)

Finally, in my third chapter, I sought to trace the history of management research, critique its development, and provide guidance for its advancement. Isaac Newton once said: *"If I have seen further, it is by standing on the shoulders of Giants"*. The idea that knowledge advances by building on the foundations laid by the ancients is not new. Building upon the shoulders of these giants requires an awareness of the newest theoretical and empirical developments in our field. In addition, the shoulders of these giants must be sturdy: our theories must be valid, and empirically testable (Rudner, 1966, p. 10) and our empirical tests must be capable of assessing causality (Antonakis, Bendahan, Jacquart, & Lalive, 2010). In this article, I conduct a co-citation analysis on all articles published in the top 50th percentile core-management journals from 1940 to 2022. This analysis allows me to summarize the development of the management literature, identify the major themes that constitute it, as well as the 20 publications that most influenced the field. I use these results to critically assess two of the most influential articles, which are representative of the broader literature, focusing on problems related to theory building and theory testing. I show that, despite being highly influential, these works contain circular theorizing and endogenous explanatory variables barring precise causal conclusions and hindering their ability to advance our knowledge of management phenomena. In all, the results of this analysis call for increased rigor in our efforts to both build and test theories.

Table of Contents

Chapter 1: Examining the use and utility of Dominance and Relative Weights Analysis

Abstract

Tyler Kleinbauer, Mikko Rönkkö, John Antonakis

Relative importance techniques are increasingly used in organizational research to assess the importance of various causes on their outcomes. These techniques are attractive because they promise to assess the importance of variables unambiguously and objectively. Regardless of their popularity, these techniques and their use have not been thoroughly scrutinized. We review the literature to see how relative importance techniques are used and critically assess if these tools are effective in addressing the problems they are used for. We find that researchers predominantly use relative importance techniques to assess the relative causal importance of variables, which these techniques are unsuitable for. We back our arguments with intuitive explanations, formal analysis of the relative importance techniques, and a Monte Carlo simulation comparing relative importance techniques against regression analysis. We show that regression provides a more accurate ranking of the relative causal importance of variables across a variety of scenarios. We conclude with practical guidelines on how to assess the relative causal importance of variables using the straightforward idea of comparable investments.

Keywords: relative importance analysis, relative weights analysis, dominance analysis, Monte Carlo Simulation, OLS, Causal importance, Comparable investments

Chapter 1: Examining the use and utility of Dominance and Relative Weights Analysis

Introduction

Is organizational culture or leadership a more important determinant of organizational performance? Such questions are common in organizational research and other social sciences but can be difficult to answer. As a consequence, organizational researchers are increasingly turning to two relative importance techniques, dominance analysis (Budescu, 1993) and relative weights analysis (Johnson, 2000), which promise to answer just this type of question. Yet, despite their increased adoption, there is little critical analysis of these techniques. Do researchers use these tools appropriately? Do the tools deliver on their promises? Our answer to both questions is "No."; the techniques do not accurately assess *causal importance*, which is the importance question that organizational researchers care about. Thus, using these techniques for causal questions can lead to false conclusions, flawed theories (Aguinis & Cronin, 2022), and ultimately, faulty policy recommendations.

We start the article by presenting a systematic review of the use of relative importance techniques in organizational research. Thereafter, we critically examine the concept of importance as conceptualized in the relative importance literature and (a) contrast it to the well-established concepts of causal and predictive importance and (b) how organizational researchers view importance, showing that there is a mismatch between what the relative importance techniques do, and what they are currently used for. We use a Monte Carlo simulation to show that the relative importance technique produces misleading answers regarding the causal importance of variables. We conclude by providing recommendations on how to assess and rank-order variables on their causal importance.

The use of relative importance techniques in organizational research

To understand the purpose for which relative importance techniques are currently used in organizational research, we searched the Web of Science for articles citing any of the seminal articles on relative importance (Azen & Budescu, 2003; Budescu, 1993; Johnson, 2000; Johnson & LeBreton, 2004; Tonidandel & LeBreton, 2011) in the “applied psychology”, “management” and “business” categories between 2016-2018. We believe this search yields a representative sample of the use of these techniques, whose popularity, across all disciplines, has been steadily increasing as can be seen in Figure 1. This search produced a list of 114 articles, of which 96 were empirical applications of the techniques and were coded further. Table 1 shows the frequency with which journals publish articles citing at least one of the five seminal articles on relative importance analysis between 2016 and 2018.

[Insert Figure 1 about here]

Table 1 shows that relative importance techniques were used mostly in the context of causal explanation, where causal terms such as *determine,* or *effect* were often employed. Frequently, explicit policy recommendations were given based on relative importance results. While some articles used predictive terms such as *predictors,* or *prediction of the outcome*, the context made it clear that these were used to reference causal concepts such as *explanatory variables* or *causal importance*. Besides determining causal importance, these techniques were also used as an attempt to address multicollinearity in the analysis stage, which was claimed to cause problems for normal regression coefficients. Relative importance was also used to assess incremental validity (Banks, McCauley, Gardner, & Guler, 2016, p. 635; Hoch, Bommer, Dulebohn, & Wu, 2018, pp. 502-503; Madrid, Totterdell, Niven, & Barros, 2016, p. 681), defined as “the unique contribution of predictors after accounting for other correlated predictors” (LeBreton, Hargis, Griepentrog, Oswald, & Ployhart, 2007, p. 472). This concept is different from the interpretation of regression coefficients as unique effects in that it refers to incremental variance. Finally, relative importance procedures were also used for other reasons like comparing the moderating effect of variables (Huang, Xu, Huang, & Liu, 2018) or validating the operationalization of the studied constructs (Beus & Whitman, 2017, p. 2153). In addition to these studies focusing on causal explanation, there were a few studies that did not make causal claims. We classified one study as predictive because the researchers were interested in predicting future credit scores (Liberati & Camillo, 2018). We also classified seven studies as descriptive because the researchers were either interested in describing the correlational structure between key variables or developing new psychometric scales.

[Insert Table 1 about here]

Our review showed that the most common use of relative importance statistics is to quantify the importance or magnitude of phenomena, making these effect sizes (Kelley & Preacher, 2012). But before we proceed to compare relative importance techniques against other statistics that could be used for the same purpose, we need to clearly define what we mean by importance.

What is importance?

Tonidandel and LeBreton (2011) define relative importance as "the contribution a variable makes to the prediction of a criterion variable by itself and in combination with other predictor variables" (p. 2). Importantly, the definition remains agnostic towards statistical and practical significance and the concept of relative importance concerns only models with direct paths between the explanatory and the outcome variables. Whereas the definition is not itself problematic, the conclusions that are drawn by researchers from its operationalization are. For example, Bernerth, Whitman, Walker, Mitchell, & Taylor (2016) ask: is interactional or procedural justice more important to increase authority figures' wellbeing? Using relative importance, they answer: "Interactional Justice climate has a stronger overall impact than Procedural Justice climate on authority figures' well-being." (p. 693). Clearly, in this study and most of the reviewed articles, relative importance is understood as a causal effect, not as a simple R^2 decomposition. Conclusions of this sort are problematic because it is possible for a variable to be highly predictively important (i.e., to greatly increase the accuracy of the prediction, i.e., the R^2, when added to the model) yet be causally irrelevant. Such a variable may then be attributed a large relative importance coefficient, which if interpreted as its causal effect will lead to faulty conclusions.

The key problem is that importance can be understood in two different ways: as causal or as predictive importance, which is confounded in the relative importance literature.

Causal importance

To define causal *importance*, we need to start from causal *relevance*. Causal relevance refers to the determinant of an effect; an event is causally relevant if the relationship between the event and the effect is not merely accidental, if that event precedes the effect, and if both the event and the effect are not caused by a common cause (Hitchcock, 2008). Causal relevance can be conceptualized using regularity, probabilistic, counterfactual, and manipulability theories of causation, which all postulate that causes make a difference in whether an effect occurs but differ on how *making a difference* is to be understood (Hitchcock, 2008, p. 324). Causal importance then refers to *the degree* to which a cause is relevant and makes a difference for an effect. If there were two relevant causes, C_1 and C_2, and C_1 is twice as powerful a determinant of an effect E than C_2, C_1 would be twice as causally important that C_2 (Achen, 1982, pp. 69-73; Bi, 2012; Bring, 1994; Darlington, 1968; Grömping, 2015). Importantly, although it is possible to compare the causal importance of two variables, the concept itself concerns just one cause and does not depend on any information about the other causes. Operationalizing the concept of causal importance requires defining the research context by answering the “where”, “who”, “what” and possibly “when” of the research. Determining the context is key because the causal importance of a variable may differ by context, for example by location or population.

Predictive importance

In prediction, researchers are interested in forecasting a likely value of a variable, based on the other variables that they have. Predictive importance can thus be defined as the degree to which a variable brings unique information to the prediction model (Grömping, 2009; Hastie, Tibshirani, & Friedman, 2009, p. 367). In this definition, variables are not considered on their own but as a part of a specific prediction model also containing other variables. Consequently, a variable’s predictive importance can vary from application to application because the other variables may change. Additionally, because in prediction the focus is on the accuracy of the forecast, the “where”, “who”, “what” and “when” questions are of lesser importance so long as the predictor are predictively important.

The difference between causal and predictive analysis

Beyond the questions of context, and the focus on a variable in isolation in causal importance, and as a part of a predictor set in predictive importance, we need to further consider the general difference between causal and predictive modeling (see Kleinberg, Ludwig, Mullainathan, & Obermeyer, 2015; Shmueli, 2010). This distinction has three important implications for research. First, in explanation the correct specification of the model is critical (Antonakis et al., 2010). Yet, in prediction, a misspecified model could perform better, as is evident from the large literature on improper linear models (Dana, Davis-Stober, & Machines, 2016; Shmueli, 2010). Second, in prediction, researchers could only be interested in the predicted value and remain agnostic towards the predictors (Hastie et al., 2009, pp. 350-352). Third, because a causally irrelevant variable can be highly predictively important, measures of predictive importance can lead to invalid policy recommendations if used for assessing causal importance. A classic example is that sales of ice cream are highly predictive of deaths by drowning (Babbie, 2014, p. 94). However, deaths by drowning are not caused by ice cream consumption. Rather, sunny weather causes both an increase in ice cream consumption and an increased risk of drowning. This does not mean that predictive models are useless for policy making. Indeed, increases in ice cream sales may be a useful signal to station more lifeguards, even if the causal mechanisms remain unknown. However, policies banning ice cream sales would be ineffective to prevent drowning regardless of how good a predictor ice cream sales are.

Statistical techniques for quantifying causal importance

We now turn to the statistical techniques. The reviewed studies mostly used relative importance techniques to quantify the magnitude of various causal phenomena. This focus on magnitude makes these statistics effect sizes (Kelley & Preacher, 2012) which they are indeed presented as in the current recommendations (LeBreton et al., 2007, p. 479; Tonidandel & LeBreton, 2011, p. 6). Therefore, we will now compare relative importance against other effect size measures shown in Table 2. These techniques can be divided into effect magnitude techniques, which are slope-based and in the same unit as the original variable, and into variance explained techniques that s are based on the square of the slope and are thus in the squared unit (Funder & Ozer, 2019). We focus on techniques that are applicable to cases where all X variables are exogenous. Extensions to more complex models (e.g. mediation) are presented in the discussion.

[Insert Table 2 about here]

Effect magnitude techniques

We will now explain four different effect magnitude techniques. Effect magnitude techniques operationalize causal importance as:

$$\textbf{Causal importance} = \textbf{Amount of the cause} \times \textbf{the strenght of the causal effect} \quad (1)$$

The different techniques use the same regression estimate (i.e., strength of effect) but differ in the "amount of the cause" multiplier used and are thus useful in different scenarios.

Raw regression coefficients

Direct comparison of the raw regression coefficients, sometimes referred to as theoretical importance (Achen, 1982, pp. 69-71), is a simple way to assess causal importance. Raw regression coefficients retain the original unit and implicitly set the "amount of the cause" of equation 1 to one. However, for this comparison to be valid, the variables must be in the same unit, and this unit must be relevant to the research question. For example, suppose one interested in losing weight asks: "Should I spend an hour running or swimming?" but has data on the calories spent per *kilometer* of swimming and those per *kilometer* of running. The units is the same, but it is not the right unit to answer the question.

Standardized regression coefficients

Another approach is to compare standardized regression coefficients, or dispersion importance (Achen, 1982), which sets the "amount of the cause" to one standard deviation (Kline, 2019, p. 151). Of course, if the variables are of a different nature, standardization does not necessarily create comparability (Greenland, Schlesselman, & Criqui, 1986; Kim & Ferree Jr, 1981; King, 1986). Consider again the swimming vs. running example, but this time we have data on the calories spent per *hour* of swimming and those per *kilometer* of running. Because the units are not comparable, standardization can be useful. This technique assumes that the variances are comparable. Consider that an inactive runner runs 10km per week (-1SD from the mean) and an active runner runs 30km per week (+1SD from the mean) and a lazy swimmer swims 2h per week (-1SD from mean) and an active swimmer swims 6h per week (+1SD from mean). For standardization to be useful, the difference in exertion (or any other relevant quantity) between the two levels (lazy vs. active) must be comparable between the two sports (Cohen, West, & Aiken, 2003, p. 154). This reveals the key weakness of standardization: it introduces a potential confound between the effect of a variable and its variance (Baguley, 2009; Greenland et al., 1986).

Many other effect size statistics are special cases of standardized regression coefficients. For example, Cohen's *d* compares the standardized mean differences between two groups, which can be derived from the coefficient of a dummy variable in regression. Other special cases are partial, semi-partial, and zero-order correlations, which are similar to standardized regression coefficients, but differ in how standardization is achieved (Stevens, 2012, pp. 75-79). Another method of standardization is range-standardization or min-max scaling which sets the range of a variable to [-1,1] or [0,1]. Because these are special cases of general techniques of quantifying dispersion importance (Achen, 1982), or techniques which are not commonly used, we will not discuss them further

Level importance

The third technique is to compare level importance which gives the effect of increasing a variable by its mean (Achen, 1982, pp. 71-73). This is similar to traditional standardization, but instead of multiplying by the standard deviation of the explanatory variable, we multiply by its mean (Hereford, Hansen, & Houle, 2004). A benefit of level importance is that when all contributions are summed, including the estimated intercept, the total equals the mean of the dependent variable. If a variable's level importance is divided by the mean of the dependent variable and multiplied by 100, it becomes a percentage representing the mean percentage increase or decrease that the variable contributes to the outcome, the sum of which totals 100%, allowing for a straightforward unconfounded comparison of variables (Achen, 1982; Kruskal, 1989). In the context of our sports example, using this approach would mean comparing these activities based on the number of calories spent by an average swimmer or runner. This technique could also be implemented using medians. Like standardization, this technique would be subject to confounding due to differences in popularity of the two sports because the means would be incomparable.

Comparable investments.

To generalize causal importance comparisons, we present the idea of comparable investments. To come up with a fair comparison in our example, we might start with the calories-per-kilometer estimates and multiply them by the number of kilometers-per-hour that a typical person can run or swim as a comparable investment. Yet, this simple strategy might be unfair because we can run directly from home, but swimming requires driving to the swimming pool. Thus, to swim for one hour, we may in fact need two or more hours. Therefore, the researcher could determine an "amount of the cause" that is deemed a comparable investment, for example, the calories spent in a two-hour window. This "amount of the cause" need not be in the same unit but must be comparable. This is the most robust method to compare variables because the scaling factor will be informed by theory and thus avoid dependence on a population quantity such as the mean, median, or standard deviation (Achen, 1982, p. 77).

Variance explained techniques

We now turn to the variance explained techniques, which are in a different class because their coefficients are in the squared unit (Funder & Ozer, 2019). Variance explained techniques are generally harder to interpret. Imagine you repeatedly play a coin-flipping game with a nickel (worth 5c) and a dime (worth 10c) (Darlington & Hayes, 2017, p. 216). If a coin comes up heads, you win the value of that coin. On each trial, you have a 25% chance of winning 15 cents, 10 cents, 5 cents, or nothing. One may ask:" What is the causal importance of the nickel and dime in determining my winnings?" To answer, we could observe the correlations between the coins and the winnings: $r_{nickel} = 0.447$ and $r_{dime} = 0.894$. The correlation of the dime is twice that of the nickel because the dime is worth twice as much as the nickel therefore the dime is twice as causally important. If we square these correlations to obtain the variance explained, we see that the nickel and the dime account for 20% and 80% of the variation in winnings. Thus, despite accurately quantifying causal importance, these measures are unintuitive and could lead to false interpretations such as claiming that the dime is four times as causally important as the nickel. Despite this weakness, variance explained measures are popular, perhaps because they can be represented on a percentage scale. We will therefore briefly explain the most common techniques.

ANOVA-based measures

The oldest group of variance explained techniques, which decompose the total variance explained in experimental designs, are based on ANOVA, a special case of regression with categorical predictors. Critically, the cells of experiments are orthogonal, which makes this decomposition possible. A common measure is R^2, called eta-squared in ANOVA. Eta-squared can be calculated for all, or a subset of variables by dividing the sum of squares of the target effects by the total sum of squares of the model. However, eta-squared is a biased measure, as it overestimates the variance explained as additional variables are added to the model, particularly at small sample sizes. Several measures which attempt to limit this upward bias exist, such as omega-squared and epsilon-squared. Epsilon-squared is noteworthy because it is identical to adjusted R^2 (Levine & Hullett, 2002). Other measures also exist such as *partial* eta, omega and epsilon squared, which quantify the variance explained by an effect after accounting for the variance explained by the other variable in the model. Surprisingly, these measures are not discussed in the relative importance literature, despite their seemingly similar goals. Other techniques can also be used to decompose R^2 like Cohen's f^2, Cohen's q or random forests (Grömping, 2009), but these are rarely used. Therefore, we will not discuss them further.

Relative importance techniques

These techniques fall into the variance explained category because they attempt to decompose the model R^2. There are extensive reviews on relative importance techniques (Bi, 2012; Johnson & LeBreton, 2004; Wei, Lu, & Song, 2015) however, we focus on the most popular techniques: dominance and relative weights analysis. When the explanatory variables are orthogonal, variance explained techniques such as eta-squared sum neatly to the model R^2. Unfortunately, this elegant result fails when variables are correlated. Dominance Analysis (Azen & Budescu, 2003; Budescu, 1993) and Relative Weights Analysis (Johnson, 2000), are presented as solutions by providing coefficients that sum to the model R^2 even if variables are correlated (Azen & Budescu, 2003; Budescu, 1993; Johnson & LeBreton, 2004; Tonidandel & LeBreton, 2011). Thus, they are claimed to be measures of effect size (LeBreton et al., 2007; Tonidandel & LeBreton, 2011). In what follows, we will focus on the fact that the coefficients of these techniques are interpreted as a variable's causal importance. The functioning of these techniques can be understood graphically with Venn diagrams. Though Venn diagrams have their limitations, they are commonly used to explain concepts of regression analysis (Kennedy, 2002).

[Insert Figure 2 about here]

Diagram A in Figure 2 shows a typical case of two explanatory variables x_1 and x_2 and a dependent variable *y*. The areas of the three circles present the total variance of each variable. The circles have four overlapping areas. The overlap between two circles represents the variance shared by two variables, which is quantified by a correlation. Areas *a* and *b* represent the unique variance explained by x_1 and x_2 respectively. This unique variation is captured as a linear relationship by regression coefficients. Area *c* is the variance shared by all three variables and area *d* is the variance shared by x_1 and x_2. The overlap between x_1 and *y* is *a*+*c*, the overlap between x_2 and *y* is *b*+*c*, and the overlap between x_1 and x_2 is *d*+*c*. The combined area *a*+*b*+*c* represents the $R^{2.}$ The objective of relative importance techniques is to decompose the R^2 into components attributable to x_1 and x_2. Because *a* should be attributed to x_1 and *b* is attributed to x_2, the decomposition question concerns the division of *c* between x_1 and x_2. In the next two sections, we provide a non-technical explanation of these techniques[1].

[1] We will not evaluate the claim that the variance of correlated variables can unambiguously be decomposed by these techniques, which has been shown to be strongly biased by sampling error and measurement unreliability (Braun, Converse, & Oswald, 2019, p. 594)

Dominance Analysis. This technique determines if a set of variables can be ranked by estimating pair-wise relationships between all variables in a regression model (Budescu, 1993). It calculates a "dominance weight" for each variable, based on the average increase to R^2 of that variable across all possible $2^p - 1$ subset models, p being the number of independent variables. The variables are then ranked according to their contribution to R^2. Suppose we have two variable, x_1 and x_2, and an outcome y, concentrating on diagram A in Figure 2, dominance analysis attempts partition the area *c* by running three regressions with three sets of predictors x_1, x_2, and both x_1 and x_2. The R^2s of these regressions are then areas *a*+*c*, *b*+*c*, and *a*+*b*+*c* respectively. To calculate the dominance weight of x_1, we first subtract the R^2 of the model with x_2 only ($R^2 = c+b$) from the R^2 of the full model ($R^2 = a+b+c$) producing $\Delta R^2 = a$. Then, we subtract the R^2 of the model with the intercept only ($R^2 = 0$), from the model with x_1 only ($R^2 = a+c$) producing $\Delta R^2 = a+c$. The dominance weight is the average of these two quantities: $a + c/2$.

Using a concrete example (R file in Appendix 2), suppose we want to study how gender identity and sex affect discrimination by sampling 2000 individuals from a large population that is half biologically female, and we measure their gender identity. Whereas gender identity and biological sex are highly correlated, they are conceptually distinct and their effects can be discriminated with sufficiently large datasets (Rönkkö & Cho, 2022). We generate our fictitious data so that discrimination depends on biological sex alone. We refer to the standardized predictor data as **X** and the standardized dependent variable as **y**[2]. Table 3 shows a subset of the data and Table 4 the correlation matrix of the data. To calculate the dominance weights and establish dominance relationships, we need to conduct the first three regressions in Table 5. To establish general dominance[3], we must determine if the average of R^2_{Gender} and $(R^2_{GenderSex} - R^2_{Sex})$ is larger than the average of R^2_{Sex} and $(R^2_{GenderSex} - R^2_{Gender})$. Because the answer is negative, as seen in

[2] We use the convention of using bold capital symbols for matrices (many variables on columns, many observations on rows) and bold lower-case symbols or vectors (one variable in a single column, many observations on rows.)

[3] Azen and Budescu (2003) establish three dominance levels, complete, conditional, and general dominance, which vary in the strictness of the dominance definition. We concentrate on general dominance, whereby a variable dominates another if the overall average additional contribution to R^2 is greater for that variable than another (Azen & Budescu, 2003, pp. 136-137), because it is the most commonly used measure.

Table 5, Sex dominates Gender (though the weights are almost indistinguishable differing by .02). Table 5 also shows that the sum of dominance weights is equal to the model R^2.

[Insert Table 3, Table 4, and Table 5 about here]

Relative Weights Analysis. This technique was developed as a less computationally intensive technique to decompose R^2 (Johnson, 2000, pp. 2-3) and approaches the problem of decomposing area *c* differently. As shown in diagram C of Figure 2, relative weights projects x_1 and x_2 as two uncorrelated variables Z_1 and Z_2. The Z variables are defined so that areas a_1+c_1 and b_2+c_2 are as large as possible, maximizing the squared correlations. The relative weights are then calculated by multiplying the squared correlations that these areas represent. The relative weight for x_1 is $(a_1+c_1)^2*a'^2 + (a_2+c_2)^2*b'^2$ and x_2 is $(b_1+c_1)^2*a'^2 + (b_2+c_2)^2*b'^2$.

Concretely, using the same dataset as before, computing relative weights requires three steps. First, the standardized original variables (**X**), are linearly transformed into new variables (**Z**) that are maximally related to the original[4], but orthogonal to each other. Thus, the first step is a principal component analysis rotated so that each component is maximally correlated with one of the original variables. Second, the dependent variable **y** is regressed on the new variables **Z**. The coefficients of this regression are called $\boldsymbol{\beta}$ in the relative weights literature. Third, the relative weights are obtained by multiplying the squared regression coefficients with the squared correlations between the **X** and **Z** variables, which are called λ in the literature. The formula for the relative weight of variable *j* is thus $\varepsilon_j = \sum_{k=1}^{p} \beta_k^2 * \lambda_{jk}^2$, *p* being the number of predictors in the model. The relative weights sum to the model R^2 (Johnson, 2000, p. 9). Highlighted in red in Table 4 are the correlation used to derive the relative weights. Table 5 shows the original regression results, the regression used for relative weight analysis, the derivation of relative weights, and a demonstration that

[4] These variables are maximally related in the least squares sense Specifically, let **X** be a matrix containing the original variable and **Z** be a matrix containing the orthogonal variables, we minimize the sum of the squared elements of $(\boldsymbol{X}-\boldsymbol{Z})$ as follows: $\boldsymbol{trace}(\boldsymbol{X}-\boldsymbol{Z})'(\boldsymbol{X}-\boldsymbol{Z}) = \boldsymbol{min}$ (Johnson, 1966). Additionally, **X**, **Z** and **y** are standardized to a mean of 0 and variance of 1

they sum to the model R^2. Table 5 also shows that the relative weights for Sex and Gender are about the same size, though Sex has a slightly higher weight; again, the difference is .02.

Criticism of the relative importance techniques

Whereas mathematical demonstrations prove that the traditional techniques presented above can accurately quantify effect size and be linked to the concept of causal importance, such mathematical proofs are still lacking with regards to the relative importance techniques. In this section, we begin by discussing problems related to collinearity, followed by logical flaws in the techniques, and finally, issues related to prediction.

Problems with the techniques when variables are correlated

Relative weights, because of their orthogonalizing properties, are presented as a solution to multicollinearity. However, it is difficult to see how, of the two relative importance techniques, relative weights specifically can solve the multicollinearity problem, when both techniques are claimed to be interchangeable (Johnson & LeBreton, 2004; Tonidandel & LeBreton, 2011). Unfortunately, neither of these techniques are a solution to the multicollinearity problem because neither can determine causality. In Table 5, we show that both the relative and dominance weights for gender and sex are the same despite the data being such that discrimination depends on sex alone. If these techniques could disentangle variable importance in the presence of collinearity, they would attribute 100% of the variation to sex, and 0% to gender, as does OLS. Indeed, when Sex and Gender are regressed together the coefficient for Sex is significant and of the correct magnitude, while gender is non-significant. Below we will show graphically, and mathematically how these techniques behave in multicollinear situations, but before, we will present the multicollinearity problem.

The multicollinearity problem

Multicollinearity refers to the degree of correlation among the independent variables in regression (Wooldridge, 2013, pp. 94-98). When multicollinearity is high, at least one variable is well explained by the others. The variance of the regression coefficients increases and the ability to say which variables have an effect is thwarted. Crucially, multicollinearity cannot be corrected in the analysis stage because it stems from the research design or context. Indeed, "claims that one can […] correct for multicollinearity are wrongheaded" (Wooldridge, 2009, p. xiv). In our previous example, because sex and gender identity are highly collinear, quantifying the effect of sex and gender identity on discrimination may be difficult when the sample contains few gender non-conforming individuals. In such a scenario, switching analysis techniques is futile. The only remedy is to gather more data, preferably on gender non-conforming individuals. From a mathematical perspective, when the OLS assumptions hold, OLS has been proven to be unbiased and efficient. Therefore, it is impossible to develop a more efficient (i.e., more precise) approach, if we require the estimates to be unbiased (i.e., correct on average).

It is possible to calculate more precise estimates, but only at the cost of systematic errors. For example, we could shrink the coefficients of gender and sex closer to zero with ridge regression (Hastie et al., 2009, pp. 61-79) which is commonly used in prediction to address the bias-variance tradeoff when the focus is on individual predictions. It is occasionally useful to introduce some bias to reduce variance, which can increase overall prediction accuracy (Shmueli, 2010). However, the deliberate introduction of bias in explanatory settings seems unproductive because studies are often aggregated into holistic assessments (e.g., in a meta-analysis) and used for policy recommendations.

Mathematical and graphical description of the behavior of these techniques when variables are correlated

In this section, we explain how each relative importance technique states that the "importance" of correlated variables is about equal when they attempt to solve the multicollinearity problem.

Consider a set of highly correlated variables x_1 and x_2 where x_1 the sole cause of y, such as in diagram B of Figure 2. If we run a regression with just x_1 or just x_2, the R^2 will be about equal in both cases because x_1 carries most of the predictive information of x_2 and vice versa. If x_1 is already in the model, then adding x_2 will hardly increase R^2 and vice versa. Therefore, regardless of whether x_1 or x_2 is the true cause, both receive a very similar dominance weight. Equations (2) and (3), which refer to diagram B of Figure 2, show how to calculate the dominance weight of x_1 and x_2. Thus, if x_1 and x_2 are collinear all the components of these equations will be almost equal, as shown in equation (4).

$$DW_{x1} = \frac{1}{2} * [R^2_{x1} + (R^2_{x1x2} - R^2_{x2})] = \frac{1}{2} * \left[(a+c) + \left((a+c) - c\right)\right] = a + \frac{c}{2} \quad (2)$$

$$DW_{x2} = \frac{1}{2} * [R^2_{x2} + (R^2_{x2x1} - R^2_{x1})] = \frac{1}{2} * \left[c + \left((a+c) - (a+c)\right)\right] = \frac{c}{2} \quad (3)$$

$$a + \frac{c}{2} = R^2_{x1} \cong R^2_{x2} = \frac{c}{2} \mid a = (R^2_{x1x2} - R^2_{x2}) \cong (R^2_{x2x1} - R^2_{x1}) = 0 \quad (4)$$

The bottom row block of Table 5 provides a concrete example, using Sex and Gender identity. This simple analysis shows two key problems of dominance analysis not covered in the prior literature. First, in the case of highly correlated variables, dominance analysis simply splits the area *c* into two halves, providing roughly equal dominance weights for both variables. Second, in cases like that of diagram B, where x_1 is the sole cause of *y*, but x_1 and x_2 are highly correlated, one could expect the dominance weight of x2 to be zero because x_2 is not a cause of *y.* Yet its dominance weight will be equal to *c*/2. For its the dominance weight to be zero, it must be uncorrelated with both x_1 and *y*, which would both areas *b* and *c* of diagram A zero.

In relative weights, the problem appears via a different mechanism. Consider diagram B in Figure 2 again where x_1 is the sole cause of *y*, but x_1 and x_2 are highly correlated. In the relative weights procedure, we construct orthogonal variables: Z_1 and Z_2, as in Diagram C, such that it be impossible for the correlation between x_1 and Z_1, and x_2 and Z_2 to be larger. We can see that relative weights analysis attempts to solve the problem of decomposing *c* (diagram A) by shifting it from the original variables to the transformed ones (diagram C). Unfortunately, this shift is not a solution. Because x_1 and x_2 are highly correlated, their correlation with Z_1 and Z_2 will be almost identical, as will the correlations between Z_1, Z_2, and *y*. Consequently, when these squared correlations are multiplied to produce the relative weights of x_1 and x_2, the results will also be almost identical because $(a_1 + c_1) \approx (b_1 + c_1)$, $(a_2 + c_2) \approx (b_2 + c_2)$ and $(a') \approx (b')$ [5]. This fact is illustrated in Table 4, if we inspect the correlation between X_{Gender} and $X_{Sex,}$ and Z_{Gender} and Z_{Sex}, and Z_{Gender} and Z_{Sex} and *y*.

As said previously, in predictive modeling, shrinkage estimators bias coefficients toward being closer to one another because the accuracy of a model is a function of bias and variance, and sometimes introducing bias increases overall prediction accuracy (James, Witten, Hastie, & Tibshirani, 2013, pp. 33-36; Kelley & Preacher, 2012; Putka, Beatty, & Reeder, 2018). However, it is unclear how the relative importance weights can be used as shrinkage estimators.

Logical and mathematical problems with relative weights

Relative weights have severe logical and mathematical flaws. First, relative weights produces orthogonal components **Z** that defy meaningful interpretation. Specifically, if **X** contain conceptually distinct variables, then **Z** confound these distinctions because they are constructed from all variables in **X**. Thus, the construct validity that applied to **X** cannot apply to **Z**, which undermines the interpretation of a model that replaces **X** by **Z** as predictors of **y**. Second, there are problems with the derivation of the link between **X** and **Z**, which was greatly debated in the literature (Green, Carroll, & DeSarbo, 1978; Johnson, 2000; Johnson & LeBreton, 2004). In an older method, Green, Carroll, and DeSarbo (1978) obtained the link between **X** and **Z** by regressing **Z** on **X**, which was criticized by J. W. Johnson (2000)[6], leading him to regress **X** on **Z** instead. Yet, considering the regression of **Z** on **X** to be flawed does not imply that the reverse is correct, and in fact, reversing the order introduces a mathematical error, which prevents relative weights from being an accurate method for decomposing R^2 (Thomas, Zumbo, Kwan, & Schweitzer, 2014). Thomas et al. (2014) show that whereas it is true that regressing **X** on **Z** yields the proportion of variance in **X** accounted for by **Z,** because the columns of **Z** are orthogonal, it is not true that one obtains the

[5] Recall that the formula to compute the relative weight of variable *j* is $\boldsymbol{\varepsilon_j} = \sum_{k=1}^{p} \boldsymbol{\beta_k^2} * \boldsymbol{\lambda_{jk}^2}$

[6] First, he claims this approach cannot accurately partition R^2 because it reintroduces the problem of correlated predictors. Second he criticizes regressing Z on X because "we are going from the orthogonal variables back to the original variables" (Johnson, 2000, p. 8).

proportion of variance in **Z** accounted for by **X,** because the columns of **X** are not orthogonal. However, the proportion of variance in **Z** accounted for by **X** is what is needed to accurately partition R^2. This difference occurs because the sum of the squared standardized regression coefficients equal the mode R^2 when the regressors are orthogonal. Thus, this results does not hold when regressing **Z** on **X**, because **X** are correlated. Failure to recognize this fact leads to the fallacy in Johnson's (2000) argument. Thomas et al. (2014, p. 334) show, in a practical empirical example, that this mathematical error leads to relative weights that can depart by as much as 12% from dominance weights. This departure could be larger in other empirical contexts. In some scenarios, both techniques differ so much that they produce different orderings of variable importance (Thomas et al., 2014). This result is alarming when both techniques are considered interchangeable (Johnson & LeBreton, 2004; Tonidandel & LeBreton, 2011).

Prediction

In the case of prediction, it is unclear how the average contribution to R^2 provided by a variable would be useful when unique contributions to R^2 are the metric of interest. In fact, Bainter, McCauley, Wager, and Losin (2020) declare that dominance, and by extension relative weights analysis, are not useful for variable selection because they rank variables within a specified model. In the case of dominance analysis, it is unclear why all possible models should be compared because usually only one or a handful of models are considered. More generally, the relative importance literature is disconnected from the modern variable selection literature (Hastie et al., 2009, pp. 57-83; James et al., 2013, pp. 203-237). Furthermore, it is unclear what advantage relative importance based variable selection would have over other techniques like forward, backward or mixed selection, comparing competing models using Mallow's C_p, AIC, BIC, or adjusted R^2 (James et al., 2013, pp. 78-79), or modern approaches like LASSO and other shrinkage techniques, which allow the inclusion of all variables and shrink the coefficients such as to obtain the best predictions (James et al., 2013, pp. 214-228). Until the superiority of relative importance-based variables selection has been demonstrated, we recommend using techniques with an established empirical and theoretical track record.

Monte Carlo simulation

To assess the performance of the relative importance techniques in causal scenarios, we ran a Monte Carlo simulation to examine the properties of the techniques and to compare them to regression analysis. Additionally, we investigated the suitability of relative weights (Johnson, 2000) to deal with multicollinearity. We designed our simulation reflects the simplest possible real-world scenario. For example, we wanted the simulation context to reflect scenarios such as explaining test scores, using age, years of education, and other variables measured without error. As such, our simulation contains an outcome as well as a varying number of explanatory variables. The rationale here is to put these three statistical techniques to the test where they should function best. Their performance should be verified in this simple case first before moving on to more complex cases.

Simulation design

The experimental factors we've manipulated, along with the values at which we manipulate them are displayed in Table 6. The parameters were chosen to reflect common research scenarios found in our systematic review and applied psychology, management, and business research. For simplicity, our variables are normally distributed with means of zero and variances of one and will be correlated according to two arbitrary correlation matrices (see Appendix 1), which are then multiplied by a constant such as to obtain variance inflation varying levels of collinearity. We vary the noise present in the model such as to obtain R^2 values of 5%, 10%, and 20%.

To check whether the relative importance procedures can disentangle correlation from causation, we generated an additional highly predictively important, yet causally irrelevant variable, *Q*. To generate *Q*, we took the fitted values of *Y* based on the population regression model, to which we added varying amounts of normally distributed random noise. Adding noise allowed us avoid perfect collinearity, and manipulate the correlation between *Q* and the fitted values of *Y*. All models were run with and without the *Q* variable to analyze its effect.

Finally, we vary the causal importance of our explanatory variables in three different ways. In one condition, all variables have the same causal importance. In another, varying causal importance, from 0 to 3. In the last condition, causal importance varies from -1 to 2 as shown in Table 6. Together, the parameters of our simulation allow us to investigate whether the relative importance techniques and regression analysis can determine the causal importance across an array of conditions.

[Insert Table 6 about here]

Following the simulation, we study the following metrics: the regression coefficients for all variables and their standard errors, the raw relative importance coefficients, and their bootstrapped 95% confidence intervals, and the model R^2. These metrics will allow us to see whether (a) the relative importance techniques can correctly quantify relative causal importance, (b) relative importance techniques can disentangle correlation and causation (c) whether relative weights analysis solves the problem of multicollinearity.

Because of its large size, the simulation was implemented on a computer cluster using R. We used the relaimpo package (Grömping, 2006) for dominance analysis and the R code provided by Tonidandel and LeBreton (2015) for the relative weight analysis. The number of experimental designs was 6'480 and we used 1000 replications per design, thus generating 6,480,000 data sets.

Simulation results

Comparison of the ordering capabilities of dominance analysis, relative weights analysis, and regression in a causal-explanatory context

To compare the accuracy of regression, dominance analysis, and relative weights analysis, we tallied the number of times each technique produced a "correct" ordering. For example, if A has a causal importance of 3, B of 2, and C of 1 in data generation, A is more important than B and C, and B is more important than C. A correct ordering occurs only if the importance order, given by the estimated coefficient, is A>B>C. In some situations, the population causal importance were equal. In those situation, we checked whether the estimated coefficients were within 5% of each other, which we deemed an appropriate cutoff for equality, taking the smallest estimate as the reference point.

[Insert Table 7 about here]

The Overall column of Table 7 shows that regression produces a correct ordering of causal importance about 15% more frequently than the relative importance techniques and all three techniques' performance is about equal whether *Q* is included or excluded from the analysis. In Table 7, we have also decomposed these results by sample size. We can see that the frequency with which regression accurately rank orders increases as sample size increases, and that the rate at which regression becomes more accurate is higher than either dominance analysis or relative weights analysis. Indeed, regression outperforms the relative importance techniques by about 10% at a sample size of 100, but by more than 20% at 50'000.

Relative importance of a causally irrelevant variable

Next, we sought to understand whether the relative importance techniques can distinguish correlation and causation. To do so, we tallied the number of times *Q*, our causally irrelevant variable, was attributed the largest raw relative importance coefficient. The results can be seen in the first two rows of Table 8. Because *Q* contained all the information of the *X* variables, a lot of variation in the outcome can be attributed to it, and increasingly so with each additional independent variable. *Q* was falsely ranked the most important variable about 15% of the time for 2 predictors and about 45% of the time for 16 predictors.

To investigate whether the estimates of *Q* were significantly different from zero, we approximated a significance test by inspecting the bootstrapped 95% confidence intervals for both techniques. We also computed the p-value from regression for the same purpose. We then compared how often the relative importance and regression coefficients for *Q* were significantly different from zero. The results of this analysis can be seen in the second-row block of Table 8. Alarmingly, this analysis shows that *Q* was almost always statistically significantly different from zero, according to the relative importance techniques. Thus, they cannot distinguish between correlation and causation because *Q* has no causal effect on the outcome. Regression analysis on the other hand does not produce false-positive results more than 5% of the time, which is expected with an alpha level of 0.05.

How do relative weights and regression analysis behave under situations of high collinearity?

Many researchers use relative weights analysis in high collinearity scenarios because standardized regression coefficients are claimed to be flawed indicators of variable importance (Johnson & LeBreton, 2004; Tonidandel & LeBreton, 2011). To investigate this claim, we ran the same analysis as in the first section, restricting our analysis to high multicollinearity scenarios (VIF=15). The results can be seen in the bottom row block of Table 7. As we can see, regression produces accurate ranking about 10% more frequently than both techniques and reacts better to sample size increases than do the relative importance techniques. We also did three Hotelling T^2 tests, one for each variation of causal importance, testing whether the means of the estimated regression coefficients were equal to a vector containing the true population causal importance (see last row of Table 6*)*. All three tests were non-significant showing the consistency and unbiasedness of regression, even under high multicollinearity. Thus, the disease of multicollinearity is simply cured by a large sample size, not relative weights (Wooldridge, 2013, p. 96).

How to assess causal importance

In this section, we briefly discuss how to assess relative causal importance. However, before operationalizing causal importance we must start by defining the context. Defining the context requires answering the “where”, “who”, “what” and possibly “when” of the research. Establishing the contexts for causal importance assessment is key to causal-explanatory research and has similarities to external validity assessment (Lucas, 2003; Lynch, 1999) in that it transcends statistical analysis and requires an understanding of the research context. After determining context, researchers can begin operationalizing causal importance. Operationalizing causal importance starts with having a research design and a statistical model that supports causal claims (Antonakis et al., 2010). We begin by discussing models where all variables are exogenous, and then move on to more complex models such as mediation models.

Causal importance in strictly exogenous models

If the model is correctly specified and causal relevance can be established, we suggest using the decision tree in Figure 3 to decide which operationalization of causal importance to implement. The first step requires determining whether the units of the variables under study are the same and/or relevant to the research question.

[Insert Figure 3 about here]

If the units are the same and relevant, researchers need to consider whether the frequencies with which the variables occurs in the population, their base rates, are important and need to be considered. For example, suppose there are two diseases, A and B, which have a 5% and 50% mortality rate respectively. If we consider the entire population, and both diseases occur with the same frequency, then comparing 5% to 50%, would provide an accurate assessment of the causal importance of these diseases. However, if Disease B were rare and A comparatively more common, infecting only 0.0002% and 2% of the population per year respectively, multiplying the regression coefficients by the base rate of occurrence would be necessary. Indeed, disease A would have a mortality rate of 0.1% and disease B 0.01%. Thus, disease A would be a more causally important cause of death than disease B.

Consider another example, drawn from our pool of reviewed studies. In her research, Tang (2016) hypothesizes that positive or negative personal turbulence (i.e., unplanned events occurring outside ones' control) may affect entrepreneur creativity. Suppose she asks: is positive or negative turbulence more causally important in explaining creativity? Assuming that the latent scales are comparable, she must consider whether the frequency of occurrence of turbulence is relevant, which depends upon the specifics of her research question: is she interested in changes in creativity per unit of turbulence, or in the overall impact of turbulence on creativity? If she is interested in the former, then comparing regression coefficients is appropriate. If she is interested in the latter, comparing level importance is appropriate. Indeed, if negative events occur twice as often as positive events, then negative events can have an overall greater impact on creativity than positive events, even if the regression coefficients are identical.

If the variables are in different units, the researcher can convert the variables to comparable and relevant units. Suppose a researcher is interested in the average hospitalization cost of diseases A and B, but has data on hospitalization rates. Answering the research question requires converting the hospitalization rate into an average cost. This can be done by obtaining data on the daily average hospitalization cost, the average days spent in hospital after the contraction of either disease and multiplying the hospitalization rate by the average daily cost and average days spent in hospital. Of course, a better conversion could be done by obtaining data on the average treatment cost of diseases A and B. If this conversion is not possible the researcher can either determine comparable investments or compare standardized regression coefficients. However, we caution the reader that standardizing confounds the effect with the variance, and does nothing to solve apples-to-oranges comparisons. One could compare a one-standard-deviation increase in genes or financial status using standardized regression coefficients; however, genes remain immutable and of a different nature than financial status.

Causal importance in more complex models

It is common that the predictors in regression are themselves causally related (Hünermund, Louw, & Rönkkö, 2022). In this case, the causal graph needs to be considered when assessing causal importance. For example, consider the following mediation model:

Studying → Learning → Exam performance

If we regress exam performance on both studying and learning, the regression coefficient of studying would be zero when controlling for learning. Can we say that studying has no effect on exam performance or that it is causally irrelevant? Of course not. In such a case, the causal importance of studying is its total effect, i.e., the sum of its direct effect and its indirect effect. Because the effect of studying is fully mediated through learning, it’s total effect is simply its indirect effect, which is obtained by multiplying the causal effect of studying on learning, by the effect of learning on exam performance.

One can then wonder whether studying is more causally important than learning? A simple answer is no because it is not studying per se that makes a difference but the learning that it produces. A more complex answer is provided by considering two causal quantities: the expected exam performance given a comparable investment of studying or learning. This comparison becomes difficult because increasing studying is straightforward, whereas increasing learning without increasing study time, which is necessary for a relevant comparison, is not. Indeed, increasing learning without study time would require increasing the rate at which studying is converted into learning.

Conclusions

In this article, we reviewed the current use of relative importance techniques in the organizational literature, showing that researchers use these techniques to determine variable importance. However, the meaning of “importance” is distinct in causal or predictive scenarios. In causal research, concepts such as *causal relevance* and *causal importance* are paramount. We examined the techniques in-depth, both analytically, and using Monte-Carlo simulations. *Overall, our results suggest abandoning the use of relative importance techniques when the focus is on causal explanations because* these techniques cannot distinguish correlation from causation *and are* outperformed by regression in all scenarios. We review several techniques capable of operationalizing causal importance, including the straightforward idea of comparable investments. Finally, as operationalizing causal importance is context-dependent, we introduce a decision tree to aid researchers in selecting the appropriate operationalization method depending on their research context. In sum, using a statistical technique that fails to determine correlation from causation when the intent is to establish causal relevance then *causal importance* is as useful as screwing in nails: one can try, but will not succeed.

References

Achen, C. H. (1982). *Interpreting and using regression* (Vol. 29): Sage.

Adams, S., Görgens-Ekermans, G., & De Kock, F. (2018). Graduate Applicant Intentions and Behavioral Beliefs. *Journal of Personnel Psychology*.

Aguinis, H., & Cronin, M. A. (2022). It's the Theory, Stupid. *Organizational Psychology Review*, 20413866221080629.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *Leadership Quarterly, 21*(6), 1086-1120. doi:10.1016/j.leaqua.2010.10.010

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods, 8*(2), 129-148. doi:10.1037/1082-989x.8.2.129

Babbie, E. R. (2014). *The practice of social research* (14th Edition ed.): Cengage learning.

Baguley, T. (2009). Standardized or simple effect size: What should be reported? , *100*(3), 603-617.

Bainter, S. A., McCauley, T. G., Wager, T., & Losin, E. A. R. (2020). Improving practices for selecting a subset of important predictors in psychology: An application to predicting pain. *Advances in Methods and Practices in Psychological Science, 3*(1), 66-80.

Banks, G. C., McCauley, K. D., Gardner, W. L., & Guler, C. E. (2016). A meta-analytic review of authentic and transformational leadership: A test for redundancy. *Leadership Quarterly, 27*(4), 634-652. doi:10.1016/j.leaqua.2016.02.006

Bernerth, J. B., Whitman, D. S., Walker, H. J., Mitchell, D. T., & Taylor, S. G. (2016). Actors have feelings too: An examination of justice climate effects on the psychological well-being of organizational authority figures. *Journal of occupational and organizational psychology, 89*(4), 693-710.

Beus, J. M., & Whitman, D. S. (2017). Almighty Dollar or Root of All Evil? Testing the Effects of Money on Workplace Behavior. *Journal of Management, 43*(7), 2147-2167. doi:10.1177/0149206314565241

Bi, J. (2012). A review of statistical methods for determination of relative importance of correlated predictors and identification of drivers of consumer liking. *Journal of Sensory Studies, 27*(2), 87-101.

Braun, M. T., Converse, P. D., & Oswald, F. L. (2019). The Accuracy of Dominance Analysis as a Metric to Assess Relative Importance: The Joint Impact of Sampling Error Variance and Measurement Unreliability. *Journal of Applied Psychology, 104*(4), 593-602. doi:10.1037/apl0000361

Bring, J. (1994). How to standardize regression coefficients. *The American Statistician, 48*(3), 209-213.

Budescu, D. V. (1993). Dominance Analysis - a New Approach to the Problem of Relative Importance of Predictors in Multiple-Regression. *Psychological Bulletin, 114*(3), 542-551. doi:Doi 10.1037//0033-2909.114.3.542

Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*: Psychology press.

Dana, J., Davis-Stober, C. P. J. M., & Machines. (2016). Rational foundations of fast and frugal heuristics: The ecological rationality of strategy selection via improper linear models. *26*(1-2), 61-86.

Darlington, R. B. (1968). Multiple Regression in Psychological Research and Practice. *Psychological Bulletin, 69*(3), 161-+. doi:DOI 10.1037/h0025471

Darlington, R. B., & Hayes, A. F. (2017). Regression analysis and linear models. *New York, NY: Guilford*.

Diefendorff, J. M., Greguras, G. J., & Fleenor, J. (2016). Perceived emotional demands–abilities fit. *Applied Psychology, 65*(1), 2-37.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156-168.

Gray, C. J., Carter, N. T., & Sears, K. L. (2017). The UWBQ-I: An adaption and validation of a measure of instigated incivility. *Journal of Business and Psychology, 32*(1), 21-39.

Green, P. E., Carroll, J. D., & DeSarbo, W. S. (1978). A new measure of predictor variable importance in multiple regression. *Journal of Marketing Research, 15*(3), 356-360.

Greenland, S., Schlesselman, J. J., & Criqui, M. H. J. A. J. o. E. (1986). The fallacy of employing standardized regression coefficients and correlations as measures of effect. *123*(2), 203-208.

Grömping, U. (2006). Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software, 17*(1). Retrieved from <Go to ISI>://WOS:000241807900001

Grömping, U. (2009). Variable importance assessment in regression: linear regression versus random forest. *63*(4), 308-319.

Grömping, U. (2015). Variable importance in regression models. *7*(2), 137-152.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction, Springer Series in Statistics. In: Springer New York.

Hereford, J., Hansen, T. F., & Houle, D. (2004). Comparing strengths of directional selection: how strong is strong? *Evolution, 58*(10), 2133-2143.

Hitchcock, C. (2008). Causation. In S. Psillos & M. Curd (Eds.), *The Routledge companion to philosophy of science* (pp. 317-326): Routledge.

Hoch, J. E., Bommer, W. H., Dulebohn, J. H., & Wu, D. Y. (2018). Do Ethical, Authentic, and Servant Leadership Explain Variance Above and Beyond Transformational Leadership? A Meta-Analysis. *Journal of Management, 44*(2), 501-529. doi:10.1177/0149206316665461

Huang, X., Xu, E., Huang, L., & Liu, W. (2018). Nonlinear Consequences of Promotive and Prohibitive Voice for Managers' Responses: The Roles of Voice Frequency and LMX. *Journal of Applied Psychology, 103*(10), 1101-1120. doi:10.1037/apl0000326

Hünermund, P., Louw, B., & Rönkkö, M. (2022). The Choice of Control Variables: How Causal Graphs Can Inform The Decision. Paper presented at the Academy of Management, Seattle.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112): Springer.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research, 35*(1), 1-19. doi:Doi 10.1207/S15327906mbr3501_1

Johnson, J. W., & LeBreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods, 7*(3), 238-257. doi:10.1177/1094428104266510

Johnson, R. M. (1966). The minimal transformation to orthonormality. *Psychometrika, 31*(1), 61-66.

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 17*(2), 137.

Kim, J.-O., & Ferree Jr, G. D. (1981). Standardization in causal analysis. *Sociological Methods Research, 10*(2), 187-210.

King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science*, 666-687.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. *105*(5), 491-495.

Kline, R. B. (2019). *Becoming a Behavioral Science Researcher: A Guide to Producing Research That Matters*: Guilford Publications.

Kruskal, W. (1989). Hooker and Yule on relative importance: A statistical detective story. *International Statistical Review/Revue Internationale de Statistique*, 83-88.

LeBreton, J. M., Hargis, M. B., Griepentrog, B., Oswald, F. L., & Ployhart, R. E. (2007). A multidimensional approach for evaluating variables in organizational research and practice. *Personnel Psychology, 60*(2), 475-498.

Levine, T. R., & Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research, 28*(4), 612-625.

Liberati, C., & Camillo, F. (2018). Personal values and credit scoring: new insights in the financial prediction. *Journal of the Operational Research Society, 69*(12), 1994-2005.

Lucas, J. W. (2003). Theory-testing, generalization, and the problem of external validity. *Journal of Sociological Theory, 21*(3), 236-253.

Lynch, J. G. (1999). Theory and external validity. *Journal of the Academy of Marketing Science, 27*(3), 367-376.

Madrid, H. P., Totterdell, P., Niven, K., & Barros, E. (2016). Leader Affective Presence and Innovation in Teams. *Journal of Applied Psychology, 101*(5), 673-686. doi:10.1037/apl0000078

Putka, D. J., Beatty, A. S., & Reeder, M. C. (2018). Modern prediction methods: New perspectives on a common problem. *Organizational Research Methods, 21*(3), 689-732.

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6-14.

Rudner, R. S. (1966). Philosophy of Social Science.

Salancik, G. R., & Pfeffer, J. (1977). Constraints on administrator discretion: The limited influence of mayors on city budgets. *Urban Affairs Quarterly, 12*, 475-498.

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, *25*(3), 289-310. doi:10.1214/10-Sts330

Shockley, K. M., Ureksoy, H., Rodopman, O. B., Poteat, L. F., & Dullaghan, T. R. (2016). Development of a new scale to measure subjective career success: A mixed-methods study. *Journal of organizational Behavior, 37*(1), 128-153.

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*: Routledge.

Tang, J. (2016). Linking personal turbulence and creative behavior: The influence of scanning and search in the entrepreneurial process. *Journal of Business Research, 69*(3), 1167-1174.

Thomas, D. R., Zumbo, B. D., Kwan, E., & Schweitzer, L. (2014). On Johnson's (2000) relative weights method for assessing variable importance: A reanalysis. *Multivariate Behavioral Research, 49*(4), 329-338.

Tonidandel, & LeBreton, J. M. (2011). Relative Importance Analysis: A Useful Supplement to Regression Analysis. *Journal of Business and Psychology, 26*(1), 1-9. doi:10.1007/s10869-010-9204-3

Tonidandel, S., & LeBreton, J. M. (2015). RWA Web: A Free, Comprehensive, Web-Based, and User-Friendly Tool for Relative Weight Analyses. *Journal of Business Psychology, 30*(2), 207-216. doi:10.1007/s10869-014-9351-z

Van Iddekinge, C. H., Aguinis, H., Mackey, J. D., & DeOrtentiis, P. S. (2018). A meta-analysis of the interactive, additive, and relative effects of cognitive ability and motivation on performance. *Journal of Management, 44*(1), 249-279

Wei, P., Lu, Z., & Song, J. (2015). Variable importance analysis: A comprehensive review. *Reliability Engineering & System Safety, 142*, 399-432.

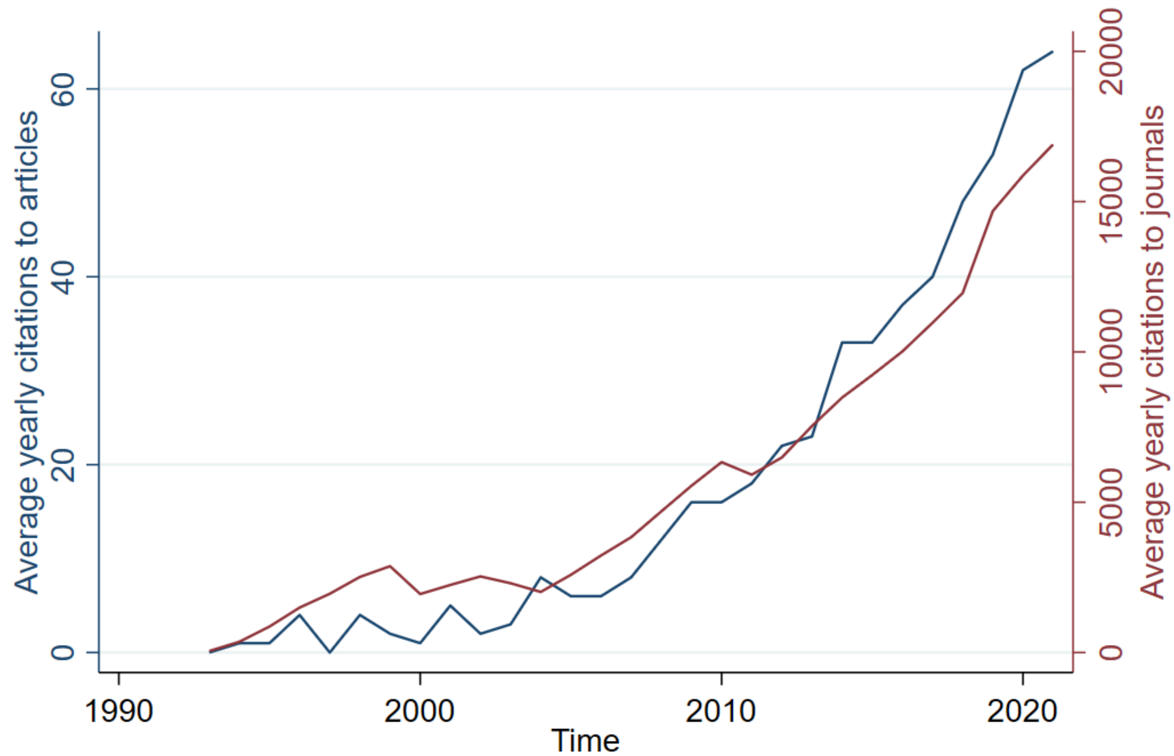
Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.): South-Western College.

.

Figures

Figure 1

Average yearly citations received by the seminal relative importance articles compared to the average citations received by articles published in those journals in which the seminal articles were published.



Note. The average yearly citation to seminal articles (Azen & Budescu, 2003; Budescu, 1993; Johnson, 2000;Johnson & LeBreton, 2004;Tonidandel & LeBreton, 2011)) are compared against the average yearly citations from all articles published in those journals in which the seminal work was published

Figure 2

Venn diagrams illustrating; (A), a typical case of two explanatory variables x_1 and x_2, and an outcome variable y; (B), of two variables x_1 and x_2, and an outcome y, where x_1 is a cause of y, and x_2 is simply correlated with x_1; and (C), two explanatory variables x_1 and x_2, orthogonal projections of these variables Z_1 and Z_2 and an outcome variable y.

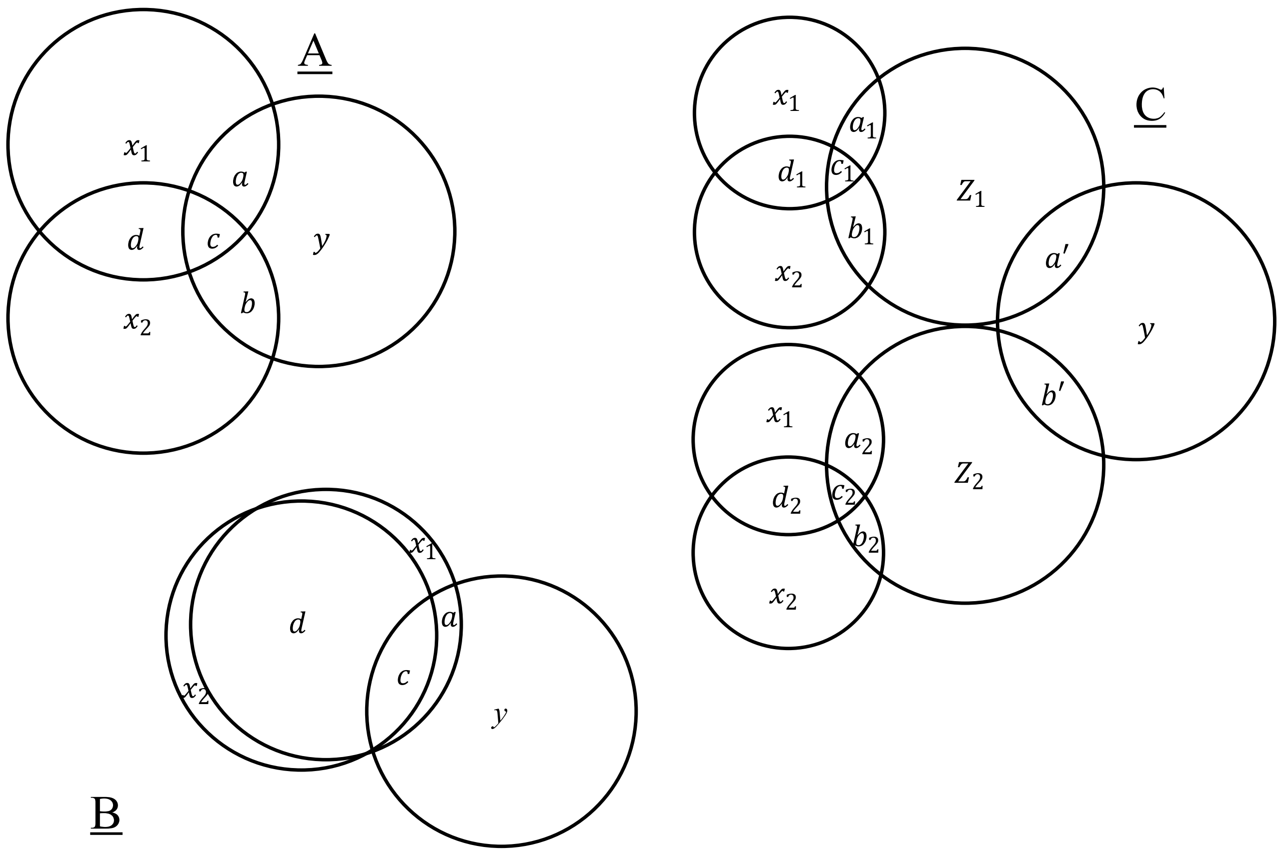
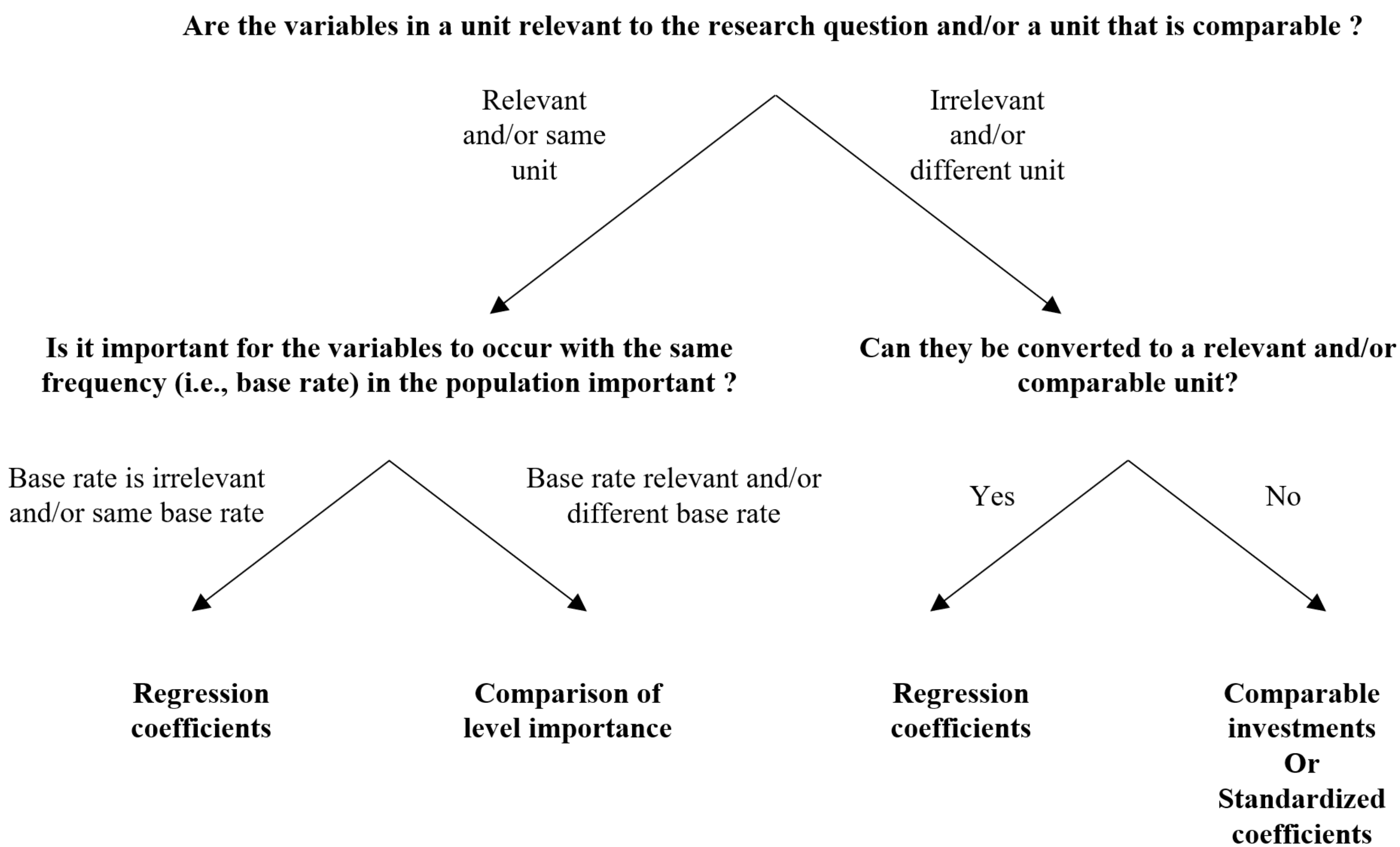


Figure 3

Decision tree for the operationalization of causal importance



Tables

Table 1

Summary of our literature review

Purpose of analysis	Reason for using relative importance	Example quotes
Causal explanation (N=88)	Determining importance (N=44)	[The] effects of ability and motivation on performance [indicated by the] relative weight for the ability-motivation interaction [are] for job performance (17.0%) […] (Van Iddekinge, Aguinis, Mackey, & DeOrtentiis, 2018, p. 265)
	Multi-collinearity (N=32)	Because […] multicollinearity […] we used relative weights analysis to determine the relative contribution of behavioral beliefs to intention to apply […] (Adams, Görgens-Ekermans, & De Kock, 2018, p. 136)
	Distinctiveness of constructs (N=4)	"[The analysis] was conducted with relative weight analyses to examine whether leader affective presence had incremental validity in predicting [dependent variables] above and beyond [other variables]." (Madrid, Totterdell, Niven, & Barros, 2016, p. 681).
	Other (N=6)	"The fact that points scored represents the strongest single determinant [*relative weights*] of individual pay in the NBA offers support for our operationalization of […] self-serving behavior." (Beus & Whitman, 2017, p. 2153)
	No justification (N=2)	
Prediction (N=1)	Multi-collinearity (N=1)	[ΔR^2 and relative weights] have been chosen to investigate two complementary aspects of the variable's importance: they consider the predictive power of a variable in isolation from and in combination with other variables in the presence of multicollinearity (Liberati & Camillo, 2018, p. 2000)
Description (N=7)	Determining importance (N=3)	[…] the relative importance of each SCSI dimension in relation to outcomes, was tested using relative weights analysis. This analysis overcomes limitations associated with multiple regression when predictors are highly correlated, as is the case with the SCSI dimensions (Shockley, Ureksoy, Rodopman, Poteat, & Dullaghan, 2016, p. 142)
	Multi-collinearity (N=3)	Dominance analysis is a useful and succinct way to parse out predictive relationships between competing correlated variables (Gray, Carter, & Sears, 2017, p. 24)
	Distinctiveness of constructs (N=1)	[…] relative weight analyses […] demonstrated that perceived ED–A fit accounted for relatively high proportions of the overall model effects [while controlling for other variables] in the regression analyses (Diefendorff, Greguras, & Fleenor, 2016, p. 15)

Table 2

Summary of the effect-size measures presented

Name of technique	Stand ardized?	Effect size dimension	Short description of the measure	The technique is useful when comparing
		Effect magnitude techniques		
Regression coefficients	No	Rate of change in original unit	Linear relationship between two variables in the original unit	Variables that are in a relevant and comparable metric
Standardized coefficients	Yes	Rate of change in standardized unit	Linear relationship between two variables in a standardized unit	Variables that are not in a comparable or relevant metrics, but confounds the causal effect with the variance
Level importance	Yes	Level in unit of the dependent variable	Contribution of the mean to level of the dependent variable	Variables that are not in comparable or relevant metrics, or have very different means
Comparable investments	Yes	Level in unit of the dependent variable	Specific net contribution to the level of the dependent variable	Variables that are in a similar metric but are incomparable.
		Variance explained techniques		
Eta-squared, Partial eta-squared, Omega Squared	Yes	Variance explained (squared unit)	Proportion of variance explained by a variable	Orthogonal variables that are not necessarily in the same unit.
Dominance analysis/ relative weights analysis	Yes	Variance explained (squared unit)	Claimed to assess proportion of variance explained by a variable	Relative importance techniques do not seem to provide additional value beyond the other techniques presented in this table.

Table 3

Fictitious scenario with two highly correlated variables: Gender and Sex, and an outcome variable Discrimination, which is caused by Sex. The first ten observations and first two gender nonconforming observations

Observation number	Raw data			Standardized data			Orthogonal predictors	
	Discrimination	Gender	Sex	$\mathbf{y}_{Discrimination}$	$\mathbf{x}_{Gender}$	$\mathbf{x}_{Sex}$	$\mathbf{z}_{Gender}$	$\mathbf{z}_{Sex}$
1	-0.89	0	0	1.21	-0.99	-0.99	-0.70	-0.70
2	-0.92	1	1	-1.24	1	1	0.71	0.71
3	1.62	0	0	0.96	-0.99	-0.99	-0.70	-0.70
4	0.51	0	0	0.00	-0.99	-0.99	-0.70	-0.70
5	0.94	1	1	0.37	1	1	0.71	0.71
6	0.70	0	0	0.16	-0.99	-0.99	-0.70	-0.70
7	0.05	0	0	-0.39	-0.99	-0.99	-0.70	-0.70
8	-1.31	0	0	-1.58	-0.99	-0.99	-0.70	-0.70
9	-1.12	1	1	-1.42	1	1	0.71	0.71
10	0.80	1	1	2.41	1	1	0.71	0.71
…	…	…	…	…	…	…	…	…
329	1.74	0	1	1.06	-0.99	1	-15.80	15.81
…	…	…	…	…	…	…	…	…
883	0.70	1	0	0.16	1	-0.99	15.81	-15.80

Note. X_{Gender} and X_{Sex} are standardized versions of Gender and Sex, thus X takes values of -.99.

Table 4

Correlations for the full fictitious scenario with two highly correlated variables: Gender and Sex, and an outcome variable Discrimination, which is caused by Sex. ***X*** *and y are the standardized raw data and* ***Z*** *is orthogonalized and standardized.*

	Mean	SD	1	2	3	4	5	6	7	8
1. Discrimination	0.51	1.16	1							
2. Gender	0.50	0.50	.44	1						
3. Sex	0.50	0.50	.44	.99	1					
4. y	0.00	1.00	1	.44	.44	1				
5. X_{Gender}	0.00	1.00	.44	1	.99	.44	1			
6. X_{Sex}	0.00	1.00	.44	.99	1	.44	.99	1		
7. Z_{Gender}	0.00	1.00	.3	.74	.67	.3	*.74*	*.67*	1	
8. Z_{Sex}	0.00	1.00	.33	.67	.74	.33	*.67*	*.74*	0	1

Note: Italicized and highlighted in red are the correlations (called lambda (λ) in the relative weights literature (Johnson, 2000)) used for the calculation of relative weights

Table 5

Regression models for discrimination, including the original regression, subset regressions for dominance analysis, and transformed regression for relative weights analysis

Regression models				
	Original	Sex	Gender	Transformed variables **Z**
Gender	-.12		1.02***	
Sex	1.15**	1.03***		
$\mathbf{Z}_{Gender}$				.30 ***
$\mathbf{Z}_{Sex}$				.33 ***
R^2	.198	.198	.196	.198
Dominance weights				
DW_{Gender}	$\frac{R^2_{Gender} + (R^2_{Original} - R^2_{Sex})}{2} = \frac{.196 + (.198 - 0.198)}{2} = .098$			
DW_{Sex}	$\frac{R^2_{Sex} + (R^2_{Original} - R^2_{Gender})}{2} = \frac{.198 + (.198 - .196)}{2} = .100$			
R^2	$DW_{Gender} + DW_{Sex} = .098 + .100 = 0.198$			
Relative weights				
RW_{Gender}	$\beta^2_{Zsex} * \lambda^2_{XGenderZSex} + \beta^2_{ZGender} * \lambda^2_{XGenderZGender} = .33^2 \times .67^2 + .30^2 \times .74^2 = .098$			
RW_{Sex}	$\beta^2_{Zsex} * \lambda^2_{XSexZsex} + \beta^2_{ZGender} * \lambda^2_{XSexZGender} = .33^2 \times .74^2 + .30^2 \times .67^2 = .100$			
R^2	$\varepsilon_{Gender} + \varepsilon_{Sex} = .0981 + .1001 = .198$			

N=2000. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Note. The relative weights are calculated by multiplying the squared correlations between **X** and **Z** (These correlation are called lambda (λ) in the relative weights literature (Johnson, 2000) and are in red in Table 4) with the squared regression coefficient of y on Z and summing the products. Intercepts are not reported because all variables are standardized.

Table 6

Design factors for the simulation

Experimental factor	Values
Sample size	100, 200, 500, 1000, 50'000
Number of explanatory variables	2, 4, 8, 16
Correlations between predictors	Matrix 1, Matrix 2 (see Appendix 1)
VIF level (collinearity)	2 (low), 7 (medium), 15 (high)
R^2 level of the estimated model	5%, 10%, 20%
Correlation of non-causal variable Q with *y*	Q not used, 0, 0.25, 0.5, 0.75, 0.9
	Variation 1 = [-1, 0, 1, 2, -1, 0, 1, 2, -1, 0, 1, 2, -1, 0, 1, 2]
Causal importance of variables[a]	Variation 2 = [0, 1, 2, 3, 0, 1, 2, 3, 0, 1, 2, 3, 0, 1, 2, 3]
	Variation 3 = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]

Note: (a) the vectors correspond to the causal importance of the explanatory variables in sequential order: [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}]

Table 7

Percentage of correct causal importance ordering of regression, dominance, and relative weights analysis, by sample size across all levels of collinearity and for high levels of collinearity only.

Technique	Sample size					
	100	200	500	1000	50'000	Overall
Analysis across all VIFs (2, 7, and 15)						
Dominance analysis	32.12%	31.94%	32.29%	33.21%	48.87%	32.39%
Relative weights analysis	32.98%	33.06%	33.98%	35.55%	51.94%	33.98%
Regression analysis	42.16%	45.06%	49.44%	53.14%	72.26%	47.46%
Analysis restricted to high collinearity (VIF = 15)						
Dominance Analysis	31.74%	31.48%	31.78%	32.58%	48.87%	31.90%
Relative weights analysis	32.82%	32.67%	33.41%	34.82%	52.65%	33.43%
Regression analysis	38.96%	41.47%	44.88%	48.21%	67.74%	43.39%

Table 8

The frequency that the causally irrelevant variable Q is given the largest coefficient by technique and number of predictors.

Technique	Number of predictors			
	2	4	8	16
Frequency with which the coefficient of Q is largest				
Dominance Analysis	15.90%	38.30%	44.30%	46.30%
Relative weights analysis	15.50%	38.50%	43.00%	44.20%
Frequency of statistical significance of the coefficient of Q				
Dominance analysis	99.33%	99.05%	98.57%	-
Relative weights analysis	99.34%	99.09%	98.50%	96.29%
Regression	5.00%	4.98%	4.99%	4.96%

Note: We did not compute bootstrapped confidence intervals for dominance analysis because of computational demands.

Appendixes

Appendix 1

Non-uniform and uniform correlation matrices used to generate the correlational structure of the data in the Monte Carlo simulation

a) Correlation matrix 1, non-uniform correlation

	X1	X1	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16
X1	1															
X2	0.5	1														
X3	0.2	0.2	1													
X4	0.4	0.2	0.2	1												
X5	0.7	0.5	0.2	0.4	1											
X6	0.5	0.5	0.2	0.2	0.5	1										
X7	0.2	0.2	0.2	0.2	0.2	0.2	1									
X8	0.4	0.2	0.2	0.4	0.4	0.2	0.2	1								
X9	0.7	0.5	0.2	0.4	0.7	0.5	0.2	0.4	1							
X10	0.5	0.5	0.2	0.2	0.5	0.5	0.2	0.2	0.5	1						
X11	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	1					
X12	0.4	0.2	0.2	0.4	0.4	0.2	0.2	0.4	0.4	0.2	0.2	1				
X13	0.7	0.5	0.2	0.4	0.7	0.5	0.2	0.4	0.7	0.5	0.2	0.4	1			
X14	0.5	0.5	0.2	0.2	0.5	0.5	0.2	0.2	0.5	0.5	0.2	0.2	0.5	1		
X15	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	1	
X16	0.4	0.2	0.2	0.4	0.4	0.2	0.2	0.4	0.4	0.2	0.2	0.4	0.4	0.2	0.2	1

b) Correlation matrix 2, uniform correlation

	X1	X1	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16
X1	1															
X2	0.5	1														
X3	0.5	0.5	1													
X4	0.5	0.5	0.5	1												
X5	0.5	0.5	0.5	0.5	1											
X6	0.5	0.5	0.5	0.5	0.5	1										
X7	0.5	0.5	0.5	0.5	0.5	0.5	1									
X8	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1								
X9	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1							
X10	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1						
X11	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1					
X12	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1				
X13	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1			
X14	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1		
X15	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	
X16	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1

Appendix 2

R code used to demonstrate the computation of relative weights

```
#
# Tonidandel, S. & LeBreton, J. M. (2014). RWA-Web -- A free, comprehensiv
e,
# web-based, and user-friendly tool for relative weight analysis. Journal
of
# Business and Psychology. doi: 10.1007/s10869-014-9351-z
#
# dominanceanalysis package


###################
# DATA GENERATION #
###################

# Remove all
rm(list=ls())

set.seed(1)
N <- 2000
sex <- sample(0:1,N, replace = TRUE)

# If 0.5% of the population are transgender or gender nonconforming  and h
alf of
# these people indicate identification to a gender opposite to their biolo
gical
# sex, then 0.25 or one in 400 indicate gender other than their sex
conforming <- runif(N)>0.0025
gender <- sex*conforming + (1-sex)*(1-conforming)

# Discrimination, determined by sex but not gender
discrimination <- 1*sex + 0*gender + rnorm(N)

# Make data frame
mydata <- data.frame(discrimination, gender, sex)

# Clean up workspace
rm(conforming, discrimination, gender, sex)


######################
# Dominance Analysis #
######################
library(dominanceanalysis)

## Warning: package 'dominanceanalysis' was built under R version 4.1.2

# dominance anaylsis
lm.da <-lm(discrimination~gender+sex, data = mydata)
da <- dominanceAnalysis(lm.da);da
```

```
##
## Dominance analysis
## Predictors: gender, sex
## Fit-indices: r2
##
## * Fit index:  r2
##        complete conditional general
## gender
## sex        gndr        gndr    gndr
##
## Average contribution:
##    sex gender
##  0.100  0.098
```

```
######################
#       RWA-Web     #
######################

# Tonidandel, S. & LeBreton, J. M. (2014). RWA-Web -- A free, comprehensiv
e,
# web-based, and user-friendly tool for relative weight analysis. Journal
of
# Business and Psychology. doi: 10.1007/s10869-014-9351-z

# https://link.springer.com/article/10.1007/s10869-014-9351-z
# https://relativeimportance.davidson.edu/Tonidandel&LeBreton(2011)_JBP_Re
lative%20Weights.pdf
# https://relativeimportance.davidson.edu/multipleregression.html

numVar<<-NCOL(mydata)
Variables<<- names(mydata)[2:numVar]

mydataCor<-cor(mydata, use="pairwise.complete.obs")
# Correlations between variables
RXX<-mydataCor[2:numVar,2:numVar]
# Correlations between X variables
RXY<-mydataCor[2:numVar,1]
# Correlations between Y and X variables

RXX.eigen<-eigen(RXX)
# Eigenvalues of the correlation of X variables
D<-diag(RXX.eigen$val)
# Create diagonal matrix with eigenvalues
delta<-sqrt(D)
# Take the square root of the eigenvalues

lambda<-RXX.eigen$vec%*%delta%*%t(RXX.eigen$vec)
# Q*LAMBDA*Q' (eqivalent to regression of X on Z)
lambdasq<-lambda^2
beta<-solve(lambda)%*%RXY
# Obtain the coefficients of the regression of y on Z
rsquare<<-sum(beta^2)

RawWgt<-lambdasq%*%beta^2
```

```r
# Mutliplying Lambda^2 by beta^2
import<-(RawWgt/rsquare)*100
# End of RWA-web code

######################
#   Alternative RWA  #
######################

# Calculation of the orthogonal variables Z
# Using the method described in (Johnson, 1966)
X <- as.matrix(scale(mydata[,2:3]))
Q <- eigen(t(X)%*%X)
D<-diag(Q$val)
delta<-sqrt(D)
delta.1<- solve(delta)
T <- Q$vec%*%delta.1%*%t(Q$vec)
Z <- scale(X%*%T)

# Calculation of the relative weights
Y <- scale(mydata[,1])
lm.yz <- lm(Y~Z)
# First regression of y on Z
coef.yz <- lm.yz$coefficients[2:3]
# Equivalent to beta in RWA-web
coef.yzsq <- coef.yz^2

lm.xz <-lm(X~Z)
# Second regression of X on Z
coef.xz <- lm.xz$coefficients[2:3,]
# Equivalent to Lambda in RWA-web
coef.xzsq <- coef.xz^2

raw.weights <- coef.xzsq%*%coef.yzsq
# Multiplying Lamdba^2 by beta^2

# End of alternative computation of Relative weights analysis


# Display that weights are the same (note: must run 2 preceeding sections)
RawWgt;raw.weights

##            [,1]
## [1,] 0.09802530
## [2,] 0.09997918

##          [,1]
## Z1 0.09802530
## Z2 0.09997918

########################################################################
#   Orthogonolizing procedure of RWA is like PCA with target rotation  #
########################################################################
```

```
# Libraries
library(psych)

## Warning: package 'psych' was built under R version 4.1.2

library(GPArotation)

# Data generation
rm(list=ls())
set.seed(1)
N <- 1000
sex <- sample(0:1,N, replace = TRUE)
conforming <- runif(N)>0.0025
gender <- sex*conforming + (1-sex)*(1-conforming)
discrimination <- 1*sex + 0*gender + rnorm(N)
mydata <- cbind(discrimination, gender, sex)
rm(conforming, discrimination, gender, sex)

# Generation of Z variables
X <- as.matrix(scale(mydata[,2:3]))
Q <- eigen(t(X)%*%X)
D<-diag(Q$val)
delta<-sqrt(D)
delta.1<- solve(delta)
T <- Q$vec%*%delta.1%*%t(Q$vec)
Z <- scale(X%*%T)

# PCA with target rotation
pcaResult <- principal(X, nfactors = 2,
                       rotate = "TargetT", Target = matrix(c(1,NA,NA,1),2,
2))

# Comparing Z to PCA results
cor(cbind(Z,pcaResult$scores))

##                                           RC1           RC2
##       1.000000e+00 -1.034477e-12 1.000000e+00 -3.957656e-06
##      -1.034477e-12  1.000000e+00 3.957655e-06  1.000000e+00
## RC1   1.000000e+00  3.957655e-06 1.000000e+00  3.753259e-15
## RC2  -3.957656e-06  1.000000e+00 3.753259e-15  1.000000e+00

summary(pcaResult$scores)

##       RC1                RC2
##  Min.   :-9.9690   Min.   :-10.0291
##  1st Qu.:-0.6738   1st Qu.: -0.7334
##  Median :-0.6738   Median : -0.7334
##  Mean   : 0.0000   Mean   :  0.0000
##  3rd Qu.: 0.7437   3rd Qu.:  0.6833
##  Max.   :10.0390   Max.   :  9.9790

summary(Z)

##        V1                 V2
##  Min.   :-9.9690   Min.   :-10.0290
```

```
##  1st Qu.:-0.6738   1st Qu.: -0.7334
##  Median :-0.6738   Median : -0.7334
##  Mean   : 0.0000   Mean   :  0.0000
##  3rd Qu.: 0.7437   3rd Qu.:  0.6833
##  Max.   :10.0390   Max.   :  9.9789
```

Chapter 2: Do leaders matter? Evidence from 56 years of U.S. governor successions

Abstract

Tyler Kleinbauer, Mikko Rönkkö, John Antonakis

Societies are facing numerous grand challenges and leaders are increasingly counted on to provide solutions. But can top-level leaders affect outcomes that unfold via various pathways? Some research streams have suggested that organizational outcomes may not be caused by, but are simply ascribed to the leader; leadership may merely be a consequence of causal attributions. We provide a rigorous test to determine whether leaders matter by exploiting a very controlled, though unusual leadership context, where leader discretion is restricted; that of U.S. state governorships. This context allows us to estimate precisely what role top-level leaders and their teams may play in determining institutional outcomes, measured on a standard metric. We quantify the "leadership effect" in a sample of 500 governors across the 50 states of the U.S. and the district of Columbia. We implement state-of-the-science methodological advances in variance decomposition on a sample of 2,985 governorship-time observations, covering the periods 1963 to 2019, to explain variance in real yearly Gross Domestic Product (GDP) growth. After having partialed out time effects (0.47 of the variance in real yearly GDP growth), we show that governors and their administrations are responsible for 2.36% of the variation in real GDP growth, while state-effects only account for 0.77% of the variation in real yearly GDP growth. Our results contradict earlier research suggesting that top-level leadership may not matter (e.g., Salancik & Pfeffer, 1977).

Keywords: Leadership, Variance decomposition, Economic growth, Leader performance, GDP, Political Leadership, Maximum likelihood, Autoregression.

Chapter 2: Do leaders matter? Evidence from 56 years of U.S. governor successions

Tyler Kleinbauer, Mikko Rönkkö, John Antonakis

Introduction

The question of leadership in society today is of capital importance. Political leaders, business leaders, and even social leaders (i.e., influencers) are all assumed to affect their relevant audiences, and thus face high levels of scrutiny. Leadership is an important question in academia too. Vroom (1976) tells us:

> There are few problems of interest to […] scientists with as much apparent relevance to the problems of society as the study of leadership. The effective functioning of social systems [to countries] is assumed to be dependent on the quality of their leadership. This assumption is reflected in our tendency to blame a football coach for a losing season or to credit a general for a military victory (p.1527).

Vroom's quote highlights the implicit assumption that leaders have an effect, and in fact most laypeople believe that "the buck stops" with leaders (Antonakis & Day, 2012). But are leaders truly responsible for the outcomes that depend upon the coordinated work of hundreds or thousands of people, or are the outcomes simply ascribed to them? Indeed, observers seem to ascribe favorable or unfavorable outcomes to the leader, regardless of the actual leader behavior (Lord, Binning, Rush, & Thomas, 1978; Rush, Thomas, & Lord, 1977). Leadership may therefore merely be the results of a causal attribution process, whereby the coordinated actions of teams generate favorable outcomes, but those outcomes are ascribed to the leader (Gemmill & Oakley, 1992; Meindl, 1995). There are conflicting empirical results supporting both perspective.

On one hand, there are multiple studies in the CEO Performance literature estimating the "leadership effect" to be responsible for about 13% to 30% of the variation in firm performance (Crossland & Hambrick, 2007; Crossland & Hambrick, 2011; Fitza, 2014; Mackey, 2008; Thomas, 1988). In fact, Quigley & Hambrick (2015) show that the CEO effect has steadily increased over time suggesting that leaders do have an effect, and increasingly so. However, these studies are challenged whereby the performance effects of randomness – a lucky turn of events leading to increased firm performance - would have been wrongly attributed to the CEO (Fitza, 2014, 2017). In these studies, Fitza shows that the CEO effect may be much closer to 5% and that this effect may be indistinguishable from chance (Fitza, 2014, 2017).

Because of its important policy implications, knowing whether leaders and their teams matter is of major importance both for leadership scholars and for society. For leadership scholars, understanding whether leaders are one of the drivers of important outcomes such as organizational performance can lead to more precise research, with important implications for leader selection and training; of course, it can also help settle the decades old debate whether leadership is merely a process of causal attribution. For society, understanding whether leadership matters can help sustain our faith in the ability of our leaders to solve crises, while highlighting the importance of those leaders that we have elected.

Our research contributes to the literature in a least three ways. First, we provide a robust test to see whether leaders and their teams matter for economic growth. We use as unique sample, U.S. state governorships from 50 US states and Washington DC from 1963 to 2019, to determine if variation in state GDP growth depends on the effect of the governor, the state or other factors. This context is one where leader discretion is restricted and where a leadership effect, if present, would take a while to unravel and to be detected at a state level; thus, this context provides for a very strong examination of whether leadership may matter. Second, we estimate this effect in a highly controlled context; by using US governors and their administrations, we compare objective, publicly observable outcome metrics (yearly GDP growth) at the state level, and thus can tease out comparable state (i.e., institution) effects versus comparable leadership (i.e., governors and their administrations) effects while accounting for common shocks to state-level GDP. Finally, we showcase a state-of-the-art method to decompose variance; we integrate the latest methodological insights in variance decomposition and use a custom maximum likelihood function in R with an autoregressive component, to estimate whether leadership matters. Importantly, our aim is to analyze only whether leadership impact economic growth, and our methodology will not allow us to investigate why, or under what conditions leadership impacts economic growth.

Attempts at quantifying the leadership effect

Leadership is a universal phenomenon seen in humans and in animals (Bass & Bass, 2009; Van Vugt, 2006). At its most fundamental level, leadership solves coordination problems like moving from point A to point B (Couzin, Krause, Franks, & Levin, 2005; King, Johnson, & Van Vugt, 2009) or coordinating on more complex tasks like sharing a public good (Antonakis, d'Adda, Weber, & Zehnder, 2021). Thus, despite poignant arguments on leadership merely being a causal attribution process, leaders believe, and are believed to have an effect on important outcomes; politicians promise to reduce crime, increase employment, and stimulate the economy; CEOs of failing companies are replaced with the expectation that they can turn a difficult situation around; leadership scholars produce thousands of research articles on the effects of leadership.

Many scholars have found evidence for the causal effect of leaders at multiple levels of analysis; in small teams (Englmaier, Grimm, Grothe, Schindler, & Schudy, 2021; Güth, Levati, Sutter, & Van Der Heijden, 2007; Levati, Sutter, & Van der Heijden, 2007), as leaders of organizations (Crossland & Hambrick, 2007; Crossland & Hambrick, 2011; Day, Rönkkö, & Antonakis, unpublished; Hambrick & Quigley, 2014; Mackey, 2008; Quigley & Hambrick, 2015; Rönkkö, Maheshwaree, & Schmidt, 2018; Thomas, 1988) and leaders of nations (Jones & Olken, 2005; Rizio & Skali, 2020). In the next two sections we will review several studies at the level of the organization and nation because this context is most like ours, in that a leader is at the helm of a large organization.

The leadership effect at the level of the organization

The CEO performance literature is large and spans multiple fields from strategic management, to economics, to corporate finance. In economics and corporate finance, several studies find that overconfident CEOs have a differential impact on corporate investments, which has downstream effects on firm profitability (Malmendier & Tate, 2005a, 2005b). Another study found that firm performance decreased following the reception of an award by the CEO, while documenting an increase in CEO activities unrelated to the role of CEO (Malmendier & Tate, 2009).

In strategic management, researchers attempt to quantify the “CEO effect”, or “leadership effect” on various metrics such as return on assets (ROA), market-to-book ratio, or cumulative abnormal returns (Crossland & Hambrick, 2007; Mackey, 2008; Quigley, Crossland, & Campbell, 2017). Some studies investigated stock market reactions to the unexpected death of the CEO (Quigley et al., 2017; Worrell & Davidson III, 1987). In the case of unexpected deaths, the researchers found that the stock market reacted in a manner consistent with the idea that CEOs have an influence over their organization. These studies are highly informative because the cause of CEO succession is exogenous. However, they do not directly measure the CEO effect.

In those studies that attempt to quantify the “leadership effect”, most apply a variance decomposition approach to determine the percentage of variation in performance attributable to the CEO. The sources of variance of the expected value of the dependent variable in CEO effect studies are generally modelled as random effects, in accordance with current recommendations (Quigley & Graffin, 2017). A key assumption in many CEO effects studies is that the effects of industry, firm, CEO and year-to-year random variations can be modeled as additive. Therefore, total performance is the sum of each variance component allowing for a straightforward decomposition. Oftentimes, several sources of variation are modeled such as an industry level effect (stable and/or changing), a firm effect (stable and/or changing), CEO effects, and then year to year random variation. Various estimation techniques, like sequential ANOVA and Multilevel models, have been used and estimate the CEO effect to be anywhere from 5% (Fitza, 2014, 2017; Thomas, 1988) to 38.5% (Hambrick & Quigley, 2014). Most studies quantify more modest effects however, ranging from about 10% to 30% (Crossland & Hambrick, 2007; Crossland & Hambrick, 2011; Fitza, 2014; Mackey, 2008); thus, it appears that the CEO is responsible for a non-trivial amount of variation in performance.

These studies suffer from certain methodological shortcomings. The greatest concern is the endogenous nature of CEO succession both with regards to the timing of the succession—which often coincides with poor performance—and with regards to the choice of successor, who is often selected to fit the needs of the organization. These issues create an endogeneity problem, that will bias any model estimates. Moreover, because timing and choice of successor are intentional this bias may be quite large. Other shortcomings also exist. First, CEOs are assumed to have a constant performance during their tenure. This assumption is unrealistic because CEOs are subject to learning effects whereby time spent within a company confers specific knowledge about its functioning leading to higher performance. Moreover, because variations in performance depend on the five parameters detailed above, omitting any of these will lead to inflated estimates of the CEO effect. This bias is called the problem of "triple confounding" by Blettner, Chaddad, and Bettis (2012), whereby disentangling persistent firm effects, CEO effects and time effects is difficult because they are collinear. Furthermore, a low number of observations can also be a concern because CEO tenures are generally short. The use of multilevel models has shown to resolve issues related to having a small number of observations, and avoid confounding the CEO effect with both the stable firm effect and random year-to-year variation (Quigley & Graffin, 2017).

Despite these methodological advances, accounting for changing firm-level difference remains a key challenge because the presence of this effect implies that firm performance correlates over time (Fitza, 2017). Therefore, methodological writings recommend the introduction of an autocorrelated error term to account for correlated firm performance (Bliese & Ployhart, 2002; Short, Ketchen Jr, Bennett, & du Toit, 2006), which allows each firm to have its own performance trajectory (Guo, 2017); in this way, one can disentangle the CEO effect from persistent firm effects which would otherwise be perfectly correlated. Two recent studies apply this learning (Day et al., unpublished; Rönkkö et al., 2018). Day, et al (unpublished) investigate the "leadership effect" in the context of scientific journal editors and their effects on the journal impact factor. They use an exponential link function to account for the multiplicative effects between journals and editors, and find that editors have a significant effect on journal impact factor. Specifically, the number of citations received by an article published by the best editors is expected to receive twice as many citations as those published under the worst editors (Day et al., unpublished). When using an additive model, they show that journal editors account for 15.7% of the variance in performance. Rönkkö, et al. (2018) investigate the effect of CEOs on ROA using an additive variance decomposition model with an autoregressive error term. They show that the CEO effect accounts for 11.5% of the variance in ROA.

The leadership effect at the level of the nation

Many researchers have studied leadership at the level of the nation. We will discuss two studies here because of their ability to make causal claims with respect to leader effectiveness, which is our aim as well. Jones and Olken (2005) studied the effectiveness of national leaders, as measured by changes in GDP growth, using a dataset of 1108 different national leaders of 130 countries from the post-world war II period to 1990. Their dataset contains autocratic and democratic countries, which has consequences for the amount of power the leader wields. By pure misfortune, several leaders died while in office, which the authors capitalized on as an exogenous source of variation. In total, 105 leaders died in office and the analysis was conducted on 57 of those deaths. The results of their analysis show that for all leaders – autocratic and democratic – one standard deviation increase in leader quality increases GDP growth rate by 1.47% which is quite a dramatic effect. If the analysis is conducted for autocrats only, they find that a one standard deviation increase in autocrat quality yields an increase in 2.1% GDP growth rate. However, the deaths of democratic leaders produce no detectable effect on growth, presumably because institutions are very strong.

In another research stream, Rizio and Skali (2020) investigate the benevolent autocrat hypothesis, whereby they tests if benevolent dictators are supposedly good for a country's economic growth and social wellbeing. Using data on the leaders of 188 countries from 1840 to 2015, they find that autocratic leaders do matter for growth. However, they find that growth-positive autocrats are found to occur only as often as chance would predict. Indeed, they find that growth-positive autocrats seem to simply be "at the right place, at the right time" and have no effect on growth. Growth-negative autocrats, on the other hand, occur much more frequently than chance levels. Together, these studies provide mixed results regarding the power of leaders to affect economic growth. It would seem like leaders can affect economic growth only if they are unconstrained in the amount of power they can wield.

Justification for our context, potential mechanisms, and hypothesis development

As is evident from the review of the literature, so far, a robust and fair test with a large sample, comparable metrics, and controls is missing. Thus, we sought out a context that could address the limitations in previous findings. As the study setting, we chose the USA, which is a very large country, having about 330 million inhabitants, and is composed of 50 states (plus the district of Columbia). In addition, the USA is adequately representative of other western democracies providing some generalizability. Each U.S. state is a semi-sovereign republic, under the umbrella of the United States Federal government. States have several rights and responsibilities such as holding elections, creating local governments, ratifying constitutional amendments and regulating intrastate commerce. Each state has its own constitution and government consisting of three branches: executive, legislative and judicial. Each states' executive branch is headed by a governor (whereas Washington D.C. is headed by a mayor).

These executives have several responsibilities and powers, including formal and informal powers (Bernick, 1979), which constitute plausible mechanism through which governors and their administrations can affect state GDP growth. Scholars distinguish up to fourteen different forms of power wielded by governors. (Bernick, 1979; Beyle, 1968; Dometrius, 1987). Theses powers can vary widely between states indeed according to some measures the least powerful governors may have half, or even a third of the powers of the most powerful governors (Beyle, 1968; Dometrius, 1987). Some of the key powers of governors are the ability to control the state budget, to appoint many officials including state judges, and to veto legislation, including vetoing line items. Thus, governors can, to varying degrees, control which activities receive funds, the political leanings of state agencies, and the direction of legislation, all of which can affect the social and economic wellbeing of a state.

The factors outline above, and others, make the US context attractive for research because the context is highly controlled. A controlled environment is the key which enables a rigorous analysis to isolate, and accurately quantify an effect, and therefore answer a research question. There are three important factors that make our context highly controlled. First, real state GDP is an objective, reliably observed, and easily comparable quantity across states. Real GDP measures the inflation adjusted monetary value of all final goods and services produced in a specific place, within a specific time period. Real GDP growth measures the change in real GDP over time. It is a broad, well established and reliable indicator of economic performance. Furthermore, it is computed according to a standard methodology making it reproducible by independent parties, and close to impossible to manipulate (in particular by the ruling administration). Second, every state is similar in its structure and organization, and has the same legal relationship to the federal government. For example, the power ceded to the states by the federal government is identical in each state. Moreover, each state is subject to the same federal laws and economic policies, especially monetary policy (e.g., interest rates, or other policies specified by the Federal Reserve). Third, each state hold gubernatorial elections at fixed intervals, usually four, making changes in leadership occur on an exogenous schedule; that is, the election trigger is not determined by the performance of the governor (unless there has been impeachment, which is very infrequent; for the timeframe of our study only two Governors have been impeached in 1988 and 2009—Arizona and Illinois respectively)[7]. All the above reasons suggest that this highly standardized context is ideal for a robust scientific test into the effect of leadership.

Given the extant CEO performance and leader succession literature, and the effects quantified therein we develop the following hypothesis.

Hypothesis: The "Governorship effect", the share of variance in GDP growth attributable to state government successions, will be non-zero.

Methodology

In this section we discuss the type of data we have collected, the context in which we will estimate a model and our modelling approach. For our study, we collected data on the district of Columbia and the 50 states and of the United States of America from 1963 to 2019. The earliest available state-level economic data is from 1963. We used various sources including the bureau of economic analysis, the US Census Bureau, and the bureau of labor statistics. Thus, we obtained the

[7]See https://www.history.com/news/us-governors-impeached-convicted-left-office for a list of impeached governors.

identity of each state governor, as well as real state GDP[8] from 1963 to 2019, yielding 2895 unique observation of real GPD on 500 governors. In some cases, the governor left or was removed from office before the end of the term. In such cases, we had two governors governing during the same year. In these cases, we attributed the economic performance of entire year to the governor that was in power longest during that year: that is, if that governor governed for more than 6 months, we counted as if that governor governed the entire year. In one case, a governor left office at exactly the mid-year mark. In this case, we attributed governorship to the exiting governor, because we believed the exiting governor would have a stronger effect on the economy than the entering governor.

The US context

In the USA, as in other locations, random year-bound shocks like energy or raw material shortages, or other shocks, may affect GDP growth. It is impossible to model every year-dependent event that could affect GPD growth; thus, the failsafe solution is to use dummy variables to capture these fixed effects of time. Furthermore, although states have similar legal structures, they are rather idiosyncratic in terms of culture, laws, geography, demography, available natural resources, and as discussed above, the power afforded to the governor, which allows for the possibility of having strong between state variability in outcomes. For instance, one state may be highly hospitable to technology companies due to its legislations, while another may contain large reserves of natural resources. These factors can naturally affect state GDP growth, and as such need to be modelled. We account for these idiosyncratic factors using a state random effect.

In addition, each state holds elections at regular intervals to elect a governor. Governors serve as chief executive officers of each of the fifty states and have considerable, yet moderated, practical powers. As such, governors can have a real effect, via policy or otherwise, on the economic health and wellbeing of a state. As such, we model a time-indexed governor random effect, which is also the variable of interest in our study. Finally, state GDP growth is also subject to various trends. For example, were a new high growth organization to settle in a state, it would contribute to the growth of that state

[8] Real GDP shows the Gross Domestic Product in constant terms, adjusted for inflation. Our data is measured in 2010 dollars.

for a significant amount of time. To account for these trends, we model a state, governor and time indexed autocorrelated error term.

Modelling Approach

The model we estimate, with our own custom likelihood function in R, includes both Governor and State effects, and a within-state autocorrelated error structure. This model consists of yearly observations (Level 1) nested in Governors (level 2), nested in States (level 3), using State real GDP growth as the dependent variable. In our analysis, we mean-center the data by year (Certo, Withers, & Semadeni, 2017; Wooldridge, 2013, pp. 484-486), which is equivalent to including year dummies, allowing us to control for macroeconomic trends. Following recommendations from the methodological literature (Bliese & Ployhart, 2002; Short et al., 2006), we specify a within state autocorrelated error term. The autocorrelation allows the model to account for an evolving state-level context and to disentangle the governor effect from these persistent state effects, which would otherwise be perfectly correlated. Because common statistical software (e,g. Stata) allows only autocorrelated errors in the lowest level groups (i.e., at the Governor level) we built a custom maximum likelihood estimator in R using the *maxLik* package (Henningsen & Toomet, 2011) to allow state-level autocorrelated errors. We obtained starting values for our custom estimator using Stata's *mixed* command. In addition, we used Stata's *mixed* command to validate our custom estimator by estimating a simpler version of our model in Stata and comparing the results, which were identical. The regression equation we estimate is displayed in Equation 5

Equation 5

Regression equation estimated by our model

$$\Delta GDP_{stg} = \underbrace{\beta_0 + \sum \beta_{1\,t} Year_t}_{fixed\ part} + \underbrace{u_s + u_{sg} + u_{sgt}}_{random\ part}$$

s = state / *t* = year / *g* = governor

We specify that GDP growth in a particular state, at a particular time, and under the leadership of a particular governor, depends on fixed and random part. $\beta_0 + \sum \beta_{1\,t} Year_t$ corresponds to the fixed part of our model where β_0 is the grand intercept and $\sum \beta_{1\,t} Year_t$ time indexed dummy variables. The random part of our model consists of the sum of three sources of variance: u_s a state indexed random intercept, u_{sg} a state and governor indexed random intercept (i.e, the "Governor Effect"), and u_{sgt} a state, governor, and time indexed autocorrelated error term. The random part of the model can be defined as the sum of three matrices, seen in Equation 6. In fact, Equation 6 is the equation we estimate with our custom likelihood function, after having accounted for the fixed part of the model.

Equation 6

Equivalent variance-covariance matrices of the random part of Equation 5

$$\boldsymbol{\Sigma_u} = \underbrace{\begin{bmatrix} \sigma_s^2 & \cdots & \sigma_s^2 \\ \vdots & \ddots & \vdots \\ \sigma_s^2 & \cdots & \sigma_s^2 \end{bmatrix}}_{state\ effect} + \underbrace{\begin{bmatrix} \begin{matrix} \sigma_{sg}^2 & \cdots & \sigma_{sg}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{sg}^2 & \cdots & \sigma_{sg}^2 \end{matrix} & & \\ & \begin{matrix} \sigma_{sg}^2 & \cdots & \sigma_{sg}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{sg}^2 & \cdots & \sigma_{sg}^2 \end{matrix} & \\ & & \begin{matrix} \sigma_{sg}^2 & \cdots & \sigma_{sg}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{sg}^2 & \cdots & \sigma_{sg}^2 \end{matrix} \end{bmatrix}}_{Governor\ effect} + \underbrace{\begin{bmatrix} \sigma_{sgt}^2 & \rho_{AR1}\sigma_{sgt}^2 & \cdots & \rho_{AR1}^{n}\sigma_{sgt}^2 \\ \rho_{AR1}\sigma_{sgt}^2 & \sigma_{sgt}^2 & \cdots & \rho_{AR1}^{n-1}\sigma_{sgt}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{AR1}^{n}\sigma_{sgt}^2 & \rho_{AR1}^{n-1}\sigma_{sgt}^2 & \cdots & \sigma_{sgt}^2 \end{bmatrix}}_{autocorrelated\ error\ variance}$$

s = state / *t* = year / *g* = governor

The first matrix contains the variance component due to the state-level random intercept u_s. Because u_s is a state-level random intercept it is constant within a state, but varies between states. It is therefore perfectly correlated over repeated observations of the state. The second matrix in Equation 6 contains a block diagonal, which represents the fact that the governor effect u_{sg} is independent between governor tenures, but constant and perfectly correlated within the same governor. The last matrix contains the error variance u_{sgt} with first order autoregression (AR1). Whereas the first two matrices were defined by one σ random effect parameter, the error variance matrix is defined by two parameters, the error variance σ_{sgt}^2 and the autoregressive parameter ρ_{AR1}. Each matrix can be estimated separately.

Results

Our analysis was conducted in two steps. First, we estimated the fixed part of

Equation 5 in Stata, before estimating the random part in R using our custom likelihood function. We began by year centering our data to account for the fixed effects of time: We regressed GDP growth on our year indicator variable, and took the predicted values. This procedure removes any variation due to macro-economic trends: indeed, using this procedure, time explains 0.47 of the variance in GDP. We use the predicted values for the rest of our analysis and the results can be seen in Table 9. Our analysis shows that state random effects account for 0.77% of the variation in state GDP growth. The state effect reflects how much states differ from the country average across all governors and time periods. The governorship random effect, the "leadership effect," accounts for 2.36% of variation in state GDP growth supporting our hypothesis. The governor effect reflects how much states differ from the country average while under the leadership of a specific governor and his or her administration. The error variance accounts for 49.88% of variation in state GDP growth. Finally, the first order autocorrelation is 0.29. Again, it is important to recall that we have removed time effects from this specification (which do account for a hefty amount of the variation).

[Insert Table 9 about here]

Discussion

In this section we discuss the results of our analysis, explain what they mean, and contrast them with other leadership literatures. We also explore the exogeneity of leadership transitions and tentatively propose that the leadership effect quantified herein can be considered causal.

On the effects we have quantified and their meaning

In our study, we have examined time, states, and governors as sources of variation in GDP growth. In Equation 5, we show that the random part of our model is represented by $\boldsymbol{u_s} + \boldsymbol{u_{sg}} + \boldsymbol{u_{sgt}}$, the state, and governor random effects, and the error variance respectively. Each of these *u* terms are point estimates of variance, however before estimating these we estimate the fixed effect of time on GDP growth. We show that 47% of the variation in state GDP can be attributed to the passage of time, which is unsurprising. Indeed, each successive year brings with it changes in a states' capability to grow. Perhaps world demand for products and serviced increase due to information one year, or perhaps

OPEC cut oil production, which lead to spikes in petrol prices (which may benefit some oil producing states though be costly for other states). Whatever the case may be, and how these shocks affect particular states is irrelevant to our specification give that we remove this source of variation from our GDP growth variable and hence focus on the random part of Equation 5.

The State effect, u_s, is state specific. It represents the difference between the average GDP growth of state *s* and the average GDP growth of the nation. This means that over the 56-year period covered by our study, states differ from the country average by 0.77%, which is the effect we have estimated. This difference could be due to various idiosyncratic state factors. For instance, some state may contain extremely fertile soil, while others may contain large reserves of natural resources, and both could generate surplus growth. This average difference of 0.77% can be seen visually in Figure 4, where we have plotted real state GDP per capita from 1963 to 2019. We have also plotted real national GDP per capita, which is represented by a black, dashed, bold line. In this figure, we can see that state growth does not differ much from national growth[9]. This result is to be expected, because the states are united in a federal republic, which redistributes resources to those state which may need economic stimulation. Indeed, economic surplus from highly urban states can be used to subsidize farming and agriculture in more rural states, which enables these states to have higher growth than otherwise possible, and "keep up" with the national average.

[Insert Figure 4 about here]

The Governorship effect, u_{sg}, represents the difference between the average GDP growth of state *s* under the leadership of governor g and his or her administration, and the average GDP growth of that state s. This measure captures any type of growth-affecting policy implemented during the tenure of a governor and his or her administration. If the growth due to this policy extends beyond the tenure of a governor, then it is captured by the error variance, which we interpret below. The governorship effect also captures organizations' future policy expectations. Indeed, organizations may choose to make investments shortly after the election of a new governor because they expect the policies implemented under this administration's leadership to be conducive to growth. We find a governor

[9] We can see that the District of Columbia, in brown, constitutes one notable exception.

effect of 2.36% which means that averaged across all governor tenures, governors and their administrations make state GDP growth deviate by 2.36% from the average state growth across all time periods.

The fact that governors and their administrations have such a pronounced effect on GDP growth, relative to the state effect, is remarkable because the United States government is specifically constituted to limit the power of elected officials. Indeed, the USA has a federalist system which distributes powers across the federal and state levels, and powers not delegated to the federal government or explicitly denied to the states are reserved to the states (U.S. Const. amend X.). Further, the USA has a tripartite system at the federal and state level, which separates the legislative, executive, and judicial branches of power granting each branch independent powers and responsibilities. The legislative branch, composed of an Upper and Lower house[10], has the sole power to create law, the executive branch governs and enforces law, and the judicial branch interprets law. These separations guarantee a system of checks and balances and prevents abuses of power. In addition to the federalist and tripartite systems, state officials must be elected which puts hopeful candidates in a competition for votes and thus imposes constraints on political vision through the median voter theorem (Downs, 1957). If elected, officials continue to face constraints on their power by other elected members which may have opposing viewpoints. Given all these constraints on governors, it is remarkable that we find support for our hypothesis and an effect as large as 2.36%.

The final random effect, accounting for both random year-to-year noise and changing state-level effect is the error variance with first-order autocorrelation, which amounts to 49.88% of the variation in state GDP growth. The error variance may seem large at first glance. However, we argue that this result is expected because the random part of our model consists of only two variables. Because we remove all year bound shocks affecting growth by year centering, the error variance does not contain any variation that would be shared by all observations such as reductions in growth due to supply-shocks. The error variance then captures every other longer-term trend which is not explained by the state or governor effects. For example, lingering economic effects due to supply shocks, such as

[10] All states except Nebraska have a bicameral legislature

bankruptcies which affect a state's economy, are captured by the error variance. Other trends such as increases in productivity due to greater computing power, more efficient communication and collaboration, and effective state-level economic policies are also captured by the error variance. Because so much of economic growth depends on these trends, which are outside of any leader's control, and orthogonal to state-specific resources, it is unsurprising for the error variance to account for 49.88% of the variation in GDP growth. Furthermore, we estimate a first-order autocorrelation of 0.29 indicating that the longer-term state-level economic trends - due to the administration's policies for example - correlate by 0.29 year-to-year. This shows that the performance trends at the state level are quite low, which could indicate that these performance trends are short lived, or that state-level performance shocks have effects far in the future, and are not captured by a one-year autocorrelation.

On the exogeneity of Leadership Change and the causality of our estimate

In this section, we argue that our study occurs within a highly controlled setting, allowing us to accurately isolate and quantify a causal governor effect. The most straightforward way of capturing a causal effect is to use an estimation procedure when the focal independent variable is exogenous. When regressing a dependent variable on an endogenous independent variable, the estimated coefficient will contain bias (Antonakis, Bendahan, Jacquart, & Lalive, 2010; Wooldridge, 2013 Chapter 2). There are multiple threats to a variables' exogeneity, however these threats usually materialize when there is a relationship between an independent variable and the error term.

In the CEO performance literature, CEO successions are not exogenous because they are generally contingent upon poor performance. Election cycles on the other hand, are an exogenous source of variation because they occur on a fixed, unchanging schedule. Most states also have term limits, which limits the time governors can stay in power even when performance is high. Additionally, governors usually complete their term duration in full, even when performance is low.

It is a common belief that economic conditions affect the outcomes of elections. However, this belief is only equivocally supported by the empirical data. Indeed, some studies do find that economic conditions affect electoral outcomes (Chubb, 1988; Fair, 1978; Happy, 1986; Kramer, 1971; Levernier, 1992; Rees, Kaufman, Eldersveld, & Freidel, 1962; Tufte, 1975), however many do not support this finding (Adams & Kenny, 1989; Arcelus & Meltzer, 1975; Erikson, 1990; Peltzman, 1987; Stigler, 1973). Four studies focus specifically on gubernatorial elections (Adams & Kenny, 1989; Chubb, 1988; Levernier, 1992; Peltzman, 1987); these show that neither the growth of state per capita personal income, nor the difference between state and federal per capita income growth significantly affect gubernatorial election outcomes. Adams and Kenny (1989) find similar results. However, Chubb (1988) finds that federal, and state per capita income growth both positively affect gubernatorial election outcomes. He also finds that national economic conditions have a stronger impact than state economic conditions. Finally, Levernier (1992) finds that per capita income growth has only a weak effect on election outcomes. Taken together, these studies suggest that in our context, changes in leadership may not be linked to performance. In fact, some suggest that voters do not view governors as being able to noticeably influence a state's economy, and do not hold them responsible nor accountable for state economic conditions (Levernier, 1992; Stein, 1990). Voters therefore evaluate gubernatorial candidates on non-economic positions suggesting that governor succession is at least exogenous with respect to performance.

Furthermore, states usually vote "red" or "blue". Indeed, in our dataset we can see that 9 out of 50 states see a change in party leadership in more than 50% of elections held. Furthermore, most states are either republican or democrat dominated. In Table 10, we can see, in those cells that are not highlighted[11], that 16 out of 50 states elect democrats or republicans about 50% of the time. The data contained in Table 10 therefore provide additional evidence that the choice of governor may be unrelated to performance. Indeed, key policy proposals are usually similar between members of the same political party.

[Insert Table 10 about here]

In addition, the link between candidate competencies and organizational needs, when compared to CEO successions, is less clear. Indeed, the entrant CEO is generally hand-picked in response to the

[11] We have highlighted the cells where either democrats or republicans govern more than 55% of the time.

organizational context whereby the CEOs competencies will fit the needs of the target firm. When it comes to governor successions, most states have party primaries (except for California, Louisiana and Washington). Thus, voters are faced with two choices – from both parties – who have been selected from a larger pool of candidates. Furthermore, although the focus of social issues may be different, both political candidates generally promise to tackle the same economic issues - such low wages for the working class - regardless of their party. The difference lies then not in the economic issue that is identified, but in the means to tackle it: some promise that wage will rise after tax cuts, and others after tax increases. This process means that voters are faced with a choice that is closer to random because which tactics will be effective at the time of voting are unclear.

The final key element allowing us to accurately isolate and quantify the leadership effect is the highly controlled context within which our study takes place. Indeed, as discussed above, state GDP growth is a standardized, objective and publicly available measure and is comparable across states and states are similar in their structure and organization. Together, this makes our context a robust candidate to isolate and accurately quantify the causal effect of leadership on economic performance.

On the accuracy of our estimate: its magnitude compared to other studies

In this section, we compare the leadership effects of our study with those from the CEO performance and national leadership literatures. Leadership can occur at multiple levels, and with varying amounts of power. Thus, comparing these effects could be considered apples-to-oranges comparisons. Nevertheless, and despite the endogenous nature of CEO successions, it may be informative to compare our estimates to those of other literatures.

Compared to governors and national leaders, CEOs may have the most latitude to affect organizational performance. Indeed, CEOs can simultaneously create, enforce, and interpret whether organizational rules are broken, thus they have access to all three "branches of power" and are relatively unconstrained. The most recent study of CEO effects, which integrates the newest methodological developments, estimates the CEO effect at 11.5% (Rönkkö et al., 2018). We propose that a governor effect of 2.36% is realistic in contrast to a CEO effect of 11.5%, because the power CEOs wield is greater than that wielded by governors, and most CEO effect studies contain positive bias. Interestingly, Rönkkö et al. (2018) find that the firm effect accounts for 21.8% of the variation in performance, whereas we find a state effect of 0.77% which may indicate that idiosyncratic organizational factors, such as culture, play a larger role in smaller institutions. Alternatively, this discrepancy may provide evidence for the power of the federal government to erase most organizational differences via redistribution.

Furthermore, we argue that, while national leaders may have more power than governors to affect average national economic growth, governors may have more power than national leader to affect the economic growth of any particular state. Indeed, as briefly discussed above, governors in the United States have more power than their national leader to affect the economic policies of their particular state. Indeed, in addition to the President, many governmental agencies such as central banks, and institutions such as Congress, play a large role in organizational performance. In addition, national leaders must govern larger bodies than governors, which may dilute performance effects. This hypothesis seems to be corroborated by Jones and Olken (2005), who find no effect of national leader succession on GDP growth in democratic countries. We therefore find it realistic that our gubernatorial estimates are non-zero. Taking these considerations into account, our study provides strong evidence for the importance of leadership on organizational performance.

Conclusion

In this research we have sought to settle the decades old debate on the causal attribution of leadership. We have provided strong evidence that leaders and their administrations do matter for economic growth by estimating this effect in a novel, highly controlled and comparable context: a sample of US governors from 50 US states and Washington DC from 1963 to 2019. We have shown that governors and their administrations are responsible for 2.36% of the variation in state GDP growth, despite multiple constraints on their power. To estimate this effect, we have used state of the art methods and integrate the latest methodological learnings to decompose the variance explained by changes in leadership.

Following our findings, and an extensive leadership literature we claim that leaders *do* matter for performance. Furthermore, against a bleak backdrop of a changing climate, mounting national debt, widening inequalities, and a worldwide pandemic, leaders have perhaps never mattered as much as now. The challenges our societies' are facing have never been greater, and are increasing in magnitude. Leadership scholars can play a role in this crisis by moving on from *whether* leaders matter on average, to determining *why* they matter and *when* they matter. In this way, we can empower leaders with sharper behavioral tools for increasingly effective policy. Indeed, we propose that future research determine the specific leader behaviors that enable and promote performance, and the conditions where those behaviors matter more or matter less.

References

Adams, J. D., & Kenny, L. W. (1989). The retention of state governors. *Public Choice, 62*(1), 1-13.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *Leadership Quarterly, 21*(6), 1086-1120. doi:10.1016/j.leaqua.2010.10.010

Antonakis, J., d'Adda, G., Weber, R. A., & Zehnder, C. (2021). Just words? Just speeches? On the economic value of charismatic leadership. *Management Science, https://doi.org/10.1287/mnsc.2021.4219*.

Antonakis, J., & Day, D. V. (2012). Leadership: Past, present, and future. The nature of leadership. Sage Publishing.

Arcelus, F., & Meltzer, A. H. (1975). The effect of aggregate economic variables on congressional elections. *American Political Science Review, 69*(4), 1232-1239.

Bass, B. M., & Bass, R. (2009). *The Bass handbook of leadership: Theory, research, and managerial applications*: Simon and Schuster.

Bernick, E. L. (1979). Gubernatorial tools: Formal vs. informal. *The Journal of Politics, 41*(2), 656-664.

Beyle, T. L. (1968). The governor's formal powers: A view from the governor's chair. *Public Administration Review*, 540-545.

Blettner, D. P., Chaddad, F. R., & Bettis, R. A. (2012). The CEO performance effect: Statistical issues and a complex fit perspective. *Strategic Management Journal, 33*(8), 986-999.

Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods, 5*(4), 362-387.

Certo, S. T., Withers, M. C., & Semadeni, M. (2017). A tale of two effects: Using longitudinal data to compare within-and between-firm effects. *Strategic Management Journal, 38*(7), 1536-1556.

Chubb, J. E. (1988). Institutions, the economy, and the dynamics of state elections. *American Political Science Review, 82*(1), 133-154.

Couzin, I. D., Krause, J., Franks, N. R., & Levin, S. A. (2005). Effective leadership and decision-making in animal groups on the move. *Nature, 433*(7025), 513-516.

Crossland, C., & Hambrick, D. C. (2007). How national systems differ in their constraints on corporate executives: a study of CEO effects in three countries. *Strategic Management Journal, 28*(8), 767-789. doi:10.1002/smj.610

Crossland, C., & Hambrick, D. C. (2011). Differences in managerial discretion across countries: how nation-level institutions affect the degree to which CEOs matter. *Strategic Management Journal, 32*(8), 797-819.

Day, D. V., Rönkkö, M., & Antonakis, J. (unpublished). Does leadership matter ? A succesion study of executive leaders in science.

Dometrius, N. C. (1987). Changing gubernatorial power: The measure vs. reality. *Western Political Quarterly, 40*(2), 319-328.

Downs, A. (1957). *An economic theory of democracy*: New York, Harper and Row.

Englmaier, F., Grimm, S., Grothe, D., Schindler, D., & Schudy, S. (2021). The Value of Leadership: Evidence from a Large-Scale Field Experiment.

Erikson, R. S. (1990). Economic conditions and the congressional vote: A review of the macrolevel evidence. *American Journal of Political Science*, 373-399.

Fair, R. C. (1978). The effect of economic events on votes for president. *The Review of Economics and Statistics*, 159-173.

Fitza, M. A. (2014). The use of variance decomposition in the investigation of CEO effects: How large must the CEO effect be to rule out chance? *Strategic Management Journal, 35*(12), 1839-1852.

Fitza, M. A. (2017). How much do CEOs really matter? Reaffirming that the CEO effect is mostly due to chance. *Strategic Management Journal, 38*(3), 802-811.

Gemmill, G., & Oakley, J. (1992). Leadership: an alienating social myth? *Human relations, 45*(2), 113-129.

Guo, G. (2017). Demystifying variance in performance: A longitudinal multilevel perspective. *Strategic Management Journal, 38*(6), 1327-1342.

Güth, W., Levati, M. V., Sutter, M., & Van Der Heijden, E. (2007). Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics, 91*(5-6), 1023-1042.

Hambrick, D. C., & Quigley, T. J. (2014). Toward more accurate contextualization of the CEO effect on firm performance. *Strategic Management Journal, 35*(4), 473-491.

Happy, J. (1986). Voter sensitivity to economic conditions: A Canadian-American comparison. *Comparative Politics, 19*(1), 45-56.

Henningsen, A., & Toomet, O. (2011). maxLik: A package for maximum likelihood estimation in R. *Computational Statistics*, *26*(3), 443-458.

James, M. (U.S. Const. amend X.). United States Constitution amendement X.

Jones, B. F., & Olken, B. A. (2005). Do leaders matter? National leadership and growth since World War II. *The Quarterly Journal of Economics, 120*(3), 835-864.

King, A. J., Johnson, D. D., & Van Vugt, M. (2009). The origins and evolution of leadership. *Current biology, 19*(19), R911-R916.

Kramer, G. H. (1971). Short-term fluctuations in US voting behavior, 1896–1964. *American Political Science Review, 65*(1), 131-143.

Levati, M. V., Sutter, M., & Van der Heijden, E. (2007). Leading by example in a public goods experiment with heterogeneity and incomplete information. *Journal of Conflict Resolution, 51*(5), 793-818.

Levernier, W. (1992). The effect of relative economic performance on the outcome of gubernational elections. *Public Choice, 74*(2), 181-190.

Lord, R. G., Binning, J. F., Rush, M. C., & Thomas, J. C. (1978). The effect of performance cues and leader behavior on questionnaire ratings of leadership behavior. *Organizational behavior and human performance, 21*(1), 27-39.

Mackey, A. (2008). The effect of CEOs on firm performance. *Strategic Management Journal, 29*(12), 1357-1367.

Malmendier, U., & Tate, G. (2005a). CEO overconfidence and corporate investment. *The journal of finance*, *60*(6), 2661-2700.

Malmendier, U., & Tate, G. (2005b). Does overconfidence affect corporate investment? CEO overconfidence measures revisited. *European financial management, 11*(5), 649-659.

Malmendier, U., & Tate, G. (2009). Superstar ceos. *The Quarterly Journal of Economics, 124*(4), 1593-1638.

Meindl, J. R. (1995). The romance of leadership as a follower-centric theory: A social constructionist approach. *The Leadership Quarterly, 6*(3), 329-341.

Peltzman, S. (1987). Economic conditions and gubernatorial elections. *The American Economic Review*, *77*(2), 293-297.

Quigley, T. J., Crossland, C., & Campbell, R. J. (2017). Shareholder perceptions of the changing impact of CEOs: Market reactions to unexpected CEO deaths, 1950–2009. *Strategic Management Journal, 38*(4), 939-949.

Quigley, T. J., & Graffin, S. D. (2017). Reaffirming the CEO effect is significant and much larger than chance: A comment on F itza (2014). *Strategic Management Journal, 38*(3), 793-801.

Quigley, T. J., & Hambrick, D. C. (2015). Has the "CEO effect" increased in recent decades? A new explanation for the great rise in America's attention to corporate leaders. *Strategic Management Journal, 36*(6), 821-830.

Rees, A., Kaufman, H., Eldersveld, S. J., & Freidel, F. (1962). The effect of economic conditions on Congressional elections 1946-1958. *The Review of Economics and Statistics*, 458-465.

Rizio, S. M., & Skali, A. (2020). How often do dictators have positive economic effects? Global evidence, 1858–2010. *The Leadership Quarterly, 31*(3), 101302.

Rönkkö, M., Maheshwaree, P., & Schmidt, J. (2018). The CEO effect and performance variation over time. *Acad Manag Proc, 1*, 16480.

Rush, M. C., Thomas, J. C., & Lord, R. G. (1977). Implicit leadership theory: A potential threat to the internal validity of leader behavior questionnaires. *Organizational behavior and human performance, 20*(1), 93-110.

Salancik, G. R., & Pfeffer, J. (1977). Constraints on administrator discretion: The limited influence of mayors on city budgets. *Urban Affairs Quarterly, 12*, 475-498.

Short, J. C., Ketchen Jr, D. J., Bennett, N., & du Toit, M. (2006). An examination of firm, industry, and time effects on performance using random coefficients modeling. *Organizational Research Methods, 9*(3), 259-284.

Stein, R. M. (1990). Economic voting for governor and US senator: the electoral consequences of federalism. *The Journal of Politics*, *52*(1), 29-53.

Stigler, G. J. (1973). General economic conditions and national elections. *The American Economic Review*, *63*(2), 160-167.

Thomas, A. B. (1988). Does leadership make a difference to organizational performance? *Administrative Science Quarterly,* 388-400.

Tufte, E. R. (1975). Determinants of the outcomes of midterm congressional elections. *American Political Science Review*, *69*(3), 812-826.

Van Vugt, M. (2006). Evolutionary origins of leadership and followership. *Personality and Social Psychology Review, 10*(4), 354-371.

Vroom, V. (1976). Leadership. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology (pp. 1527–1551). Chicago, IL: Rand McNally.

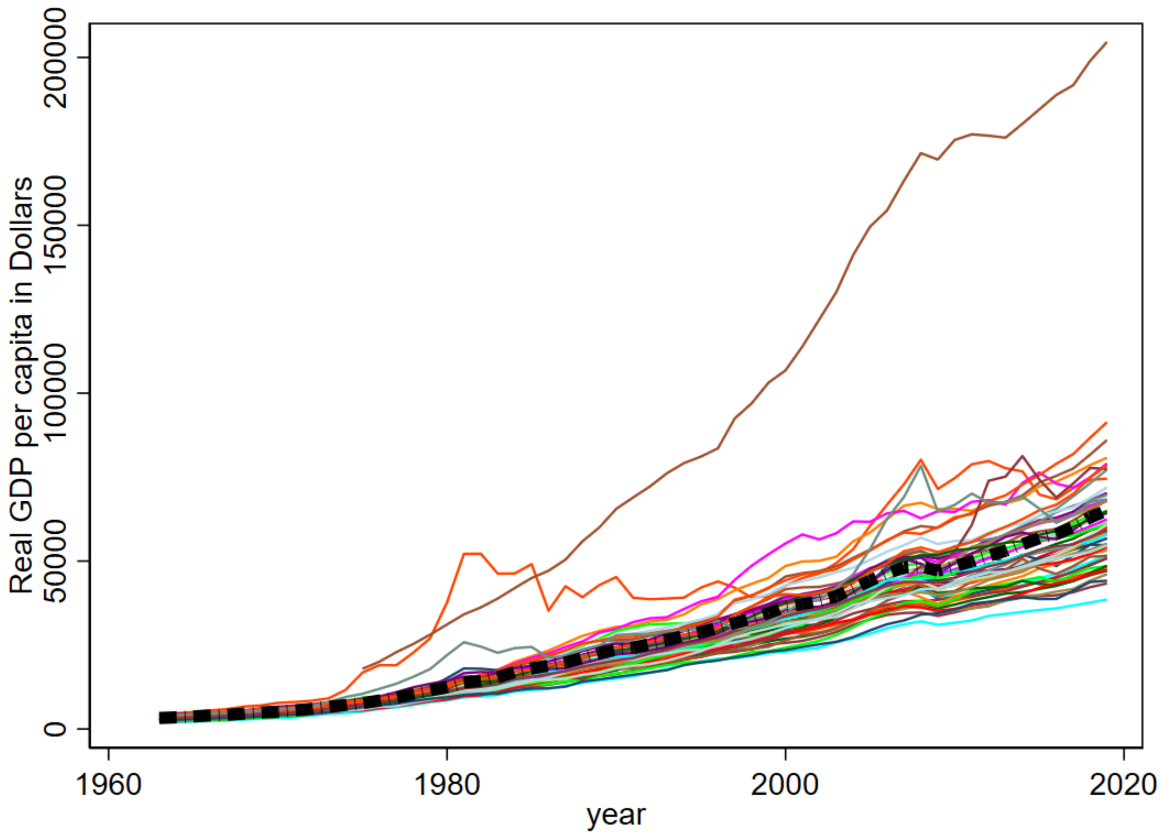
Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.): South-Western College.

Worrell, D. L., & Davidson III, W. N. (1987). The effect of CEO succession on stockholder wealth in large firms following the death of the predecessor. *Journal of Management, 13*(3), 509-515.

Figures

Figure 4

Plot of state real GDP per capita from 1963 to 2019, with national GDP per capita in black, bold and dashed.



Tables

Table 9

Mixed effects regression of state GDP growth (after accounting for macroeconomic trends)

	Model estimates	Percentage of variance
Fixed Effects		
Intercept	1.0700000	
	(0.0048100)	
Random Effects		
State	0.000008692	0.77%
	(0.0000031588)	
Governor	0.000026765	2.36%
	(0.000005989)	
Error variance	0.0005671	49.88%
	(0.000009646)	
AR1 (within state)	0.29900000	
	(0.02090000)	
Log likelihood	5775.70800000	

Note: AR1 is in correlation metric.

Table 10

Frequency with which a democrat or republican governor is in power, and frequency of switches in governing party relative to number of elections held in that state.

State	Democrat	Republican	Other	Party Switches	Number of elections	Frequency of party switch
Alabama	52.63%	47.37%	-	5	15	33.30%
Alaska	42.11%	40.35%	17.55%	10	14	71.40%
Arizona	40.35%	59.65%	-	8	16	50.00%
Arkansas	63.16%	36.84%	-	7	21	33.30%
California	45.61%	54.39	-	6	15	40.00%
Colorado	64.91%	35.09%	-	3	15	20.00%
Connecticut	57.89%	35.09%	7.02%	5	15	33.30%
Delaware	64.91%	35.09%	-	4	14	28.60%
Florida	49.12%	50.88%	-	5	15	33.30%
Georgia	70.18%	29.82%	-	1	15	6.70%
Hawaii	85.96%	14.04%	-	2	14	14.30%
Idaho	42.11%	57.89%	-	2	15	13.30%
Illinois	40.35%	59.65%	-	6	15	40.00%
Indiana	38.60%	61.40%	-	3	14	21.40%
Iowa	31.58%	68.42%	-	3	18	16.70%
Kansas	50.88%	49.12%	-	9	18	50.00%
Kentucky	78.95%	21.05%	-	5	15	33.30%
Louisiana	57.89%	35.09%	7.02%	8	14	57.10%
Maine	43.86%	35.09%	21.05%	8	15	53.30%
Maryland	80.70%	19.30%	-	5	15	33.30%
Massachusetts	45.61%	54.39%	-	5	16	31.30%
Michigan	29.82%	70.18%	-	5	16	31.30%
Minnesota	15.79%	42.11%	42.11%	8	15	53.30%
Mississippi	57.89%	42.11%	-	3	14	21.40%
Missouri	59.65%	40.35%	-	7	14	50.00%
Montana	61.40%	38.60%	-	3	14	21.40%
Nebraska	42.11%	57.89%	-	7	16	43.80%
Nevada	50.88%	49.12%	-	6	15	40.00%
New Hampshire	49.12%	50.88%	-	7	29	24.10%
New Jersey	50.88%	49.12%	-	8	14	57.10%
New Mexico	57.89%	42.12%	-	8	17	47.10%
New York	57.89%	42.11%	-	3	15	20.00%
North Carolina	71.93%	28.07%	-	6	14	42.90%
North Dakota	45.61%	54.39%	-	3	15	20.00%
Ohio	28.07%	71.93%	-	6	15	40.00%
Oklahoma	49.12%	50.88%	-	6	15	40.00%
Oregon	64.91%	35.09%	-	3	15	20.00%
Pennsylvania	50.88%	49.12%	-	7	15	46.70%
Rhodes Island	50.88%	49.12%	-	5	23	21.70%
South Carolina	42.11%	57.89%	-	5	15	33.30%
South Dakota	14.04%	85.96%	-	2	18	11.10%
Tennessee	49.12%	50.88%	-	7	15	46.70%
Texas	42.11%	57.89%	-	5	18	27.80%
Utah	35.09%	64.91%	-	2	14	14.30%
Vermont	57.89%	42.11%	-	9	29	31.00%
Virginia	57.89%	42.11%	-	6	14	42.90%
Washington	71.93%	28.07%	-	4	14	28.60%
Washington D.C.	100%	-	-	0	12	0.00%
West Virginia	66.67%	33.33%	-	7	14	50.00%
Wisconsin	40.35%	59.65%	-	8	17	47.10%
Wyoming	49.12%	50.88%	-	4	15	26.70%
MEANS	**52.12**	**46.01**	**1.87**	-	-	**33.55%**[a]

Notes: The mean of the frequency of party switches is significantly less than 50%. 95% CI [0.29;0.38]

Chapter 3: A Frankensteinian patchwork: a co-citation network analysis and critique of management articles from 1940 to 2022

Tyler R. Kleinbauer – University of Lausanne

Abstract

Isaac Newton once said: *"If I have seen further, it is by standing on the shoulders of Giants"*. The idea that knowledge advances by building on the foundations laid by the ancients is not new. Building upon the shoulders of these giants requires an awareness of the newest theoretical and empirical developments in our field. In addition, the shoulders of these giants must be sturdy: our theories must be valid, and empirically testable (Rudner, 1966, p. 10) and our empirical tests must be capable of assessing causality (Antonakis, Bendahan, Jacquart, & Lalive, 2010). In this article, I conduct a co-citation analysis on all articles published in the top 50th percentile core-management journals from 1940 to 2022. This analysis allows me to summarize the development of the management literature, identify the major themes that constitute it, as well as the 20 publications that most influenced the field. I use these results to critically assess two of the most influential articles, which are representative of the broader literature, focusing on problems related to theory building and theory testing. I show that, despite being highly influential, these works contain circular theorizing and endogenous explanatory variables barring precise causal conclusions and hindering their ability to advance our knowledge of management phenomena. In all, the results of this analysis call for increased rigor in our efforts to both build and test theories.

Chapter 3: A Frankensteinian patchwork: a co-citation network analysis and critique of management articles from 1940 to 2022

Tyler R. Kleinbauer – University of Lausanne

Introduction

"If I have seen further, it is by standing on the shoulders of Giants"

– *Isaac Newton*

The idea that knowledge advances by building on the foundations laid by the ancients is not new, and stretches back at least until the 12th century when Bernard of Chartres said: *nanos gigantum humeris insidentes*, or "Dwarves mounted on the shoulders of giants". This idea has become trite and even serves as the tagline for Google Scholar. Nevertheless, it remains at the core of the scientific enterprise: we cite others to build on the work of those who came before us. Standing on the shoulders of giants means building upon the cutting-edge, which requires an awareness of the newest theoretical and empirical developments in our fields. The field of management has grown immensely in recent years (see Figure 5) and reading summaries has become a useful way of keeping up. Indeed, 11'071 articles were published between the years 2018 and 2020 in the top core-management journals, compared to 437 between 1940 and 1942.

[Insert Figure 5 about here]

Narrative reviews and meta-analyses are useful qualitative and quantitative ways of centralizing and summarizing current knowledge. However, those methods focus on specific topics. As of now, there are no broad, systematic, and objective reviews of the field as a whole, and its evolution over time. A useful way of visualizing the entire field, and its evolution, is a co-citation network analysis. A co-citation analysis (McCain, 1990) uses co-citation counts to construct measures of similarity between documents, authors, or journals. Co-citation is defined as the frequency with which two units are cited together (Small, 1973). Importantly, the assumption behind a co-citation analysis is that the more two items are cited together, the more likely it is that their content is related (Zupic & Čater, 2015). Co-citation analysis can then be enhanced with network visualization. The benefit of using co-citation analysis is that it reveals both the most influential publications, and thematic clusters allowing to uncover seminal publications and knowledge foundations of a field or research program (Donthu, Kumar, Mukherjee, Pandey, & Lim, 2021). When examined over time, co-citation networks can help detect paradigm shifts (Kuhn, 1970), and changes in schools of thought, shedding light on the major themes underpinning the intellectual structure and development of a field (Donthu et al., 2021; Pasadeos, Phelps, & Kim, 1998).

Co-citation analysis thus groups cited references into thematic clusters. These clusters are the giants whose shoulders we wish to stand on. To stand, however, the shoulders of these giants must be sturdy: our theories must be valid, and empirically testable (Rudner, 1966, p. 10) and our empirical tests must be capable of assessing causality (Antonakis et al., 2010). Yet, some argue that many organizational-level theories are so vague they can never be empirically tested (Bacharach, 1989). Others argue that, because of current publication pressures, authors generate theory for theories' sake, which may then be inconsistent and of questionable validity (Aguinis & Vandenberg, 2014; Antonakis, 2017; Ferris, Hochwarter, & Buckley, 2012; Tourish, 2020; Van de Ven, 1989). Consequently, our theories grow in a disjunct, ad-hoc and unsystematic manner hindering the very way science and understanding progress (Antonakis, 2017).

However, a good theory is valuable because it can provide "revelatory insight" (Corley & Gioia, 2011), which allows us to: "see profoundly, imaginatively, unconventionally into phenomena we thought we understood" (Mintzberg, 2005, p. 361). Claims of inconsistent and invalid theories seem to be borne out by the facts because most theoretical statements are never empirically tested (Edwards, 2010; Edwards, Berry, & Kay, 2014; Kacmar & Whitfield, 2000). Furthermore, those rare theoretical propositions that do get empirically tested often contain endogeneity (Antonakis, Bastardoz, & Rönkkö, 2021; Fischer, Dietz, & Antonakis, 2017; Hamilton & Nickerson, 2003), which biases estimates and hinders causal interpretability (Antonakis et al., 2010). The consequence of poor theorizing on the one hand, and lack of and/or faulty empirical testing on the other is lack of replicability, lack of impact in terms of influencing policy, and ultimately a potential drying up of resources dedicated to researching management phenomena.

Because of the need for a broad, systematic, and objective review of our field, both in terms of its structure and evolution, I performed a co-citation analysis and network visualization of the field of management from 1940 to 2022. To understand the forces that shaped this literature, I used this analysis to identify the most influential research themes in each period, and the contribution of different disciplines like economics and psychology to management research. Furthermore, because management is in dire need of valid, empirically testable theory (Aguinis & Cronin, 2022; Cronin, Stouten, & van Knippenberg, 2021), and sound, causally interpretable empirical work (Antonakis et al., 2021; Fischer et al., 2017; Hamilton & Nickerson, 2003), I identified and summarized the 20 most influential works over the 1940-2022 period, and critically assess the theoretical and empirical soundness of two exemplars of the broader literature. Together, I believe my summary and critique can help identify where the field has been, and where it should go in the future to make new scientific advances.

Literature Review

In this section, I will provide a brief history of the development of the field of management, give a definition of management based on its historical purpose, discuss two modes of contribution to the scientific literature - theory building and theory testing – and provide a brief summary of desirable properties, and threats to the validity of both modes of contribution.

History and definition of management

Research in the social sciences has a long history. Disciplines and bodies of scholarly thought such as economics, psychology, and sociology originated in the late 18th and early 19th centuries (Agarwal & Hoetker, 2007) whereas management, as a discipline, is much younger and arose in the late 19th century from a practical need for skilled business managers. Indeed, the first business school was founded in 1881 at the University of Pennsylvania by Joseph Wharton who thought "it would be sensible for young men to learn to do something useful in College and not merely how to conjugate Latin verbs or strum upon the mandolin" (Silk, 1964, p. 421). Mr. Wharton thought that American college education was failing young Americans and sought to create a college that would fit those attending "for the actual duties of life". His college would teach them of the necessity "of organizing under single leaders or employers great amounts of capital and great numbers of laborers and of maintaining discipline among the latter" such as to produce a class of men "ready to assume leadership and command in practical and civil affairs" (Silk, 1964, p. 422). The purpose of management education is thus to understand and master the functioning and command of organizations. From then on, the number of business schools increased rapidly, particularly in the post-World War II era, prompted in part by a growing demand for professional managers (Agarwal & Hoetker, 2007; Bennis & O'Toole, 2005). However, in an influential report published in 1959, Gordon and Howell noted that: "[The] business literature is not, in general, characterized by challenging hypotheses, well developed conceptual frameworks, the use of sophisticated research techniques, penetrating analysis, the use of evidence drawn from relevant underlying disciplines - or very significant conclusions (Gordon & Howell, 1959, p. 379). Following this critique, business schools sought to improve their rigor and by the end of the 20th century, many business schools began offering a curriculum of academic distinction (Bennis & O'Toole, 2005).

Despite these efforts, significant progress remains to be made, both in terms of the theoretical, and empirical work conducted in our field. For example, Miner (1984) found that out of 32 reviewed theories, only 11 were rated high in scientific validity. 19 years later, Miner (2003) shows that out of 73 reviewed management theories, only 25 were rated high in scientific validity. These studies show severe deficits in the quality of management theorizing. Similarly, several scholars show that a vast majority of theoretical propositions are never empirically tested (Edwards, 2010; Edwards et al., 2014; Kacmar & Whitfield, 2000). In terms of those propositions that are empirically tested, several scholars have found that about 80 to 90% of quantitative management articles contain at least one of the 7 threats to validity illustrated in Table 11, which biases the estimated coefficients and hinders causal interpretability (Antonakis, Bastardoz, Liu, & Schriesheim, 2014; Antonakis et al., 2021; Antonakis et al., 2010; Fischer et al., 2017). These threats to validity will be discussed in a lower section. Because scientific advances in management can be categorized either as theory building or theory testing

(Colquitt & Zapata-Phelan, 2007), which both rest on different modes of reasoning (Mantere & Ketokivi, 2013), I will discuss both in turn in the next section.

Two approaches to advancing management research: theory building and theory testing

Scholarly work can be divided into two broad approaches: theory building and theory testing (Colquitt & Zapata-Phelan, 2007). Each approach is undergirded by a different mode of reasoning. Consider the classic example by Peirce (1878):

1. All the beans in this bag are white (The Rule)
2. These beans are from this bag (The Explanation)
3. These beans are white (The Observation)

When building theory, scholars are interested in inferring the rule (1) from the explanation (2) and the observation (3), which is why some advocate for building theory from case studies (Eisenhardt, 1989). Because theory building begins with the observation and an inductive approach, it is often associated with but is not limited to, qualitative research. When testing theory, scholars adopt a deductive approach whereby they deduce the observation (3) from the rule (1) and the explanation (2). Because of the many existing estimation techniques and statistical tests, which enable the quantification of relationships and the testing of hypotheses in a relatively objective way (see chapter 4 of Wooldridge, 2013), quantitative approaches are particularly well suited to theory testing. Quantitative approaches, given a properly specified statistical model, allow researchers to tentatively reject the absence of a causal effect between two variables, which is a prerequisite for accurate and ethical policy recommendations (Antonakis et al., 2010). Another form of reasoning, abductive reasoning, also exists whereby the researcher infers the explanation (2) from the rule (1) and the observation (3) (Mantere & Ketokivi, 2013). This form of reasoning is useful when selecting which theoretical explanation fits the data best. Next, I delineate desirable properties of both theoretical and empirical work.

Desirable properties in theory building

Theory is a statement of a set of falsifiable causal relations among concepts, including lawlike generalizations, within a set of boundary assumptions and constraints. The purpose of theory is to parsimoniously organize and clearly communicate these relations (Bacharach, 1989; Rudner, 1966, p. 10; Whetten, 1989). In other words, theory is the answer to the question: “what is going on?” (Aguinis & Cronin, 2022). Therefore, theory goes beyond a mere description of raw data, typology, and metaphor (Bacharach, 1989), that is, in addition to explaining the “what”, a theory must explain the “when”, “how”, and most importantly the “why” (Whetten, 1989). In this way, theories can both explain why something has happened, but also predict if it will happen again (Aguinis & Cronin, 2022; Hunt, 1991, p. 149).

Building on the work of many scholars (Bacharach, 1989; Hunt, 1991; McKelvey, 1997; Priem & Butler, 2001; Rudner, 1966; Whetten, 1989), three criteria can be used to evaluate whether a set of statements can be considered a scientific theory: these statements must be lawlike in that they (1) are generalized conditionals, (2) have empirical content, and (3) have nomic necessity. First, generalized conditionals are if/then statements, such as: *if* leaders display increased amounts of individualized consideration, *then* follower job satisfaction will increase. An if/then structure is necessary for a statement to qualify as a generalized conditional. Second, the empirical content criterion addresses the semantics and logic of a theory. It helps separate purely analytical statements from synthetic statements. Analytical statements, because of their either/or form, or because of the way the terms of the statement are defined do not necessitate empirical investigation to determine their veracity. For example, statements such as, “it is either raining or not raining”, or “outgoing individuals will score higher on extraversion” are true by definition. Synthetic statements, such as “as new employees are socialized in the organization, their performance will increase” can only be known to be true after investigation (Hunt, 1991). Third, nomic necessity is the criterion that demands that “the occurrence of some phenomenon must be associated with some other phenomenon; the relationship cannot be, simply, by chance” (Hunt, 1991, p. 111). Taken together, these three criteria, which by themselves are necessary but not sufficient, allow the assessment of the validity of a set of statements. If all three criteria are satisfied, a set of statements can be considered scientifically valid.

Desirable properties in theory testing

Theory testing, the purpose of which is establishing causal relations between a set of variables, requires following the hypothetico-deductive model, whereby theoretical propositions are formulated into testable hypotheses (Bacharach, 1989). The researcher then attempts to falsify these hypotheses with observation (Sankey, 2013). Failing to falsify a hypothesis makes it accrue “money in the bank” or a “good track record” (Meehl, 1990). Our confidence in the truth of the causal relationship between the hypothesized variables should increase proportionally to the amount of “money in the bank” this relationship has accrued (Meehl, 1990). The gold standard for establishing causality is the experiment, particularly the randomized field experiment (Antonakis et al., 2010; Lonati, Quiroga, Zehnder, & Antonakis, 2018). However, conducting randomized field experiments may not always be feasible. Therefore, other methods, like instrumental variable estimation, also exist that permit the establishment of causal relationships in the absence of exogenous manipulation (Antonakis et al., 2010; Ketokivi & McIntosh, 2017).

Consider two variables, *x*, and *y*. If the relation between *x* and *y* is due, in part, to other reasons, then *x* is endogenous (Antonakis et al., 2010; Ketokivi & McIntosh, 2017). Technically, in the context of OLS regression, endogeneity refers to the situation where an independent variable correlates with the structural error term in a model (Wooldridge, 2002, 2013). Failure to exogenously manipulate a variable or to implement the appropriate statistical method results in endogeneity. Antonakis et al. (2010) identified 7 categories of threats to the validity of quantitative analyses. These threats are reproduced in Table 11, If one of these threats is present, endogeneity bias becomes a concern.

[Insert Table 11 about here]

Endogeneity bias is severely problematic because it can bias regression coefficients upward, or downward and can even change their sign, hindering the causal interpretability of these estimates

(Antonakis et al., 2010; Ketokivi & McIntosh, 2017). In a later section, I will critique two highly influential scientific articles using the criteria discussed in the two previous sections. These articles are representative exemplars of the broader management literature. Before however, I will discuss the methods I have used to conduct a broad and objective co-citation analysis of the management literature.

Methods

Data selection

A crucial part of my analysis was selecting a core set of documents to be analyzed. I wanted my analysis to identify seminal articles and research programs in the field of management, and how these evolved. The challenge was therefore to select a broad enough base of articles such that they are representative of the field and its evolution, while also limiting their scope to a manageable size. I, therefore, limited my analysis to articles published in the most influential core management journals, from 1940 to 2022[12] for three reasons. First, I assume that seminal articles have a higher probability of being published in a top journal. Second, I assume that seminal articles have a higher probability of being published in a generalist journal. Third, few management-centric journals existed before the 1950 to 1960 period (see Table 11), indicating that management as a discipline was not fully developed yet. However, because co-citation analysis analyzes citation patterns, influential books and articles published outside of the selected journals, or selected periods would still be included in the analysis.

To select journals that fit these criteria, I downloaded a list of all journals categorized as "management" on the Web Of Science SSCI, along with their respective 2020 impact factors. I then eliminated the bottom 50th percentile of management journals according to their impact factor, before manually inspecting the list to remove all non-core-management journals[13]. For example, I eliminated journals such as the Journal of Service Management, Tourism management, or MIS quarterly. The final list of journals that I used as a basis for downloading scientific articles can be seen in Table 12.

[12] The data were downloaded on the 23.03.2022. The exact search can be seen using the following link: https://www.webofscience.com/wos/woscc/summary/bda7b37c-0f94-4027-94c6-4b192df628e8-2065f0ed/relevance/1. The time range must be adapted to those specified in Table 13

[13] Non-core-management journals are management journals dedicated to a specialist domain

[Insert Table 12 about here]

After finalizing the list of source journals, I proceeded to search the Web Of Science Core Collection for all articles published in the journals listed in Table 12. I searched for all articles published in these journals from 1940 to 2022, separated in the time ranges listed in Table 13. I chose these time ranges for three main reasons. First, I wanted the width of these time ranges to reflect the speed at which science advanced, which I operationalized as the number of published scientific articles in a given period. In Table 13, we can see that, excluding the 2021-2022 period, the number of published articles oscillates between 5200 and 11'000. Second, because time also plays a role in the advancement of knowledge - ideas take time to assimilate - I wanted to avoid redundant analyses by creating too narrow time bins. Thus, I chose a minimum time bin of 3 years. Third, I wanted to avoid computational issues which arise when too many references are included in the co-citation network. The chosen time ranges allow us to see the evolution of research over time. I then downloaded the full citation and all cited references for each article.

[Insert Table 13 about here]

Data analysis strategy

One of the key challenges was generating a co-citation map that accurately, yet parsimoniously represents the field over the chosen periods. Because I was interested in the main research paradigms, I chose to limit my analysis to approximately the top 1% most influential articles per period. I operationalized the influence of an article by its citation count which has certain drawbacks because citations are a noisy indicator of influence. To limit my analysis, I adjusted the minimum number of times a reference must be cited to be included in the co-citation map (see Table 13). The number, and nature of the clusters did not substantially change if I chose a higher or lower cutoff than the one displayed.

I used the VOSViewer software to analyze the bibliometric data (Van Eck & Waltman, 2010). The software allows the direct upload of Web Of Science data. I used the software to perform a co-citation analysis and a network visualization. Together, these analyses create a visual map allowing one to easily see clusters of highly co-cited works. Sometimes, the Web Of Science data was incomplete, whereby in place of the reference name, [no title captured] was displayed. Because VOSviewer is unable to distinguish these cases from real references, I manually excluded [no title captured] from the analysis. After the identification of the co-citation network for each selected period, I downloaded the title and abstract for the five most highly cited articles for each cluster, yielding a list of 456 abstracts and titles, of which 225 were unique. I used these titles and abstracts to identify the major theme of each cluster.

Clustering and mapping algorithms

Three types of maps can be distinguished in bibliometric research: distance-based maps, graph-based maps and time-based maps. VOSViewer creates distance-based maps, which are maps in which the distance between two items reflects the strength of the relation between the items. Several algorithms can be used to create such maps and VOSViewer uses the "visualization of similarities" (i.e., VOS) technique which has distinct advantages over other popular techniques (Van Eck, Waltman, Dekker, & Van Den Berg, 2010), and is a consistent approach borne out of the same underlying principles and assumptions (Waltman, Van Eck, & Noyons, 2010).

VOSViewer constructs a map in three steps. First a similarity matrix is calculated based on the co-occurrence matrix. To obtain the similarity matrix, the co-occurrence matrix must be normalized by correcting the latter for differences in the total number of occurrences or co-occurrences of items. To do this it uses a similarity measure known as the association strength (Van Eck & Waltman, 2007; van Eck, Waltman, van den Berg, & Kaymak, 2006). Second, a map is constructed by applying the VOS mapping technique to the similarity matrix. This techniques constructs a map such that the distance between two items reflects their similarity: a small distance reflects high similarity, whereas a large distance reflects low similarity. The idea of the VOS mapping technique is to minimize the weighted sum of the squared Euclidean distances between all pairs of items. Third, the map is translated, rotated and reflected to ensure that the solution to the minimization problem of step two is globally optimal. For the mathematical details of the VOS technique see Van Eck and Waltman (2010) and Waltman et al. (2010).

Results of the Network Analysis

In this analysis, I identified 92 clusters over the 16 time periods. The results of the analysis are presented in Table 14 and Table 15. Table 14 presents the results of the analysis from the years 1940 to 1999 and Table 15 from the years 2000 to 2022. The tables display the major theme of each cluster, for each period, which I've identified using the five most highly cited articles for each cluster. Thus, I used 456 abstracts and titles, of which 225 were unique to extract the extract and identify the major theme. The five most highly cited articles per cluster, per period, and the co-citation map can be seen in the appendix. To the right of the major theme, I have added the proportion of cited references attributed to each research area. This proportion indicates the activity of each major research theme in that period. For a list of the top five most cited articles per cluster, which were used to identify the major theme of each cluster, see Appendix 3.

[Insert Table 14 about here]

[Insert Table 15 about here]

Contribution of disciplines to the selected references

A researcher generally adopts a disciplinary perspective – psychological, sociological, or economic, when conducting research. To understand the evolution, and contribution of these various disciplinary perspectives to the chosen literature, I categorized the 225 unique references into 1 of 8 categories: Mathematics, Anthropology, Sociology, Management, Economics, Psychology, Statistics, and Biology. These eight categories were selected inductively as I was parsing through the 225 references. To categorize each reference, I used the definition of each discipline displayed in Table 16.

[Insert Table 16 about here]

[Insert Figure 6 about here]

In Figure 6, we see that the three most dominant perspectives used to conduct research in our selected sample are Management, Psychology, and Sociology. We can see that between the years 1940 to about 1990, psychology was the dominant perspective used to study management phenomena.

Nevertheless, management was already an influential perspective in the 1940s and is the dominant perspective in 2022. During the period between 1940 and 1970, sociology only had a weak influence on the selected sample. However, it became an important perspective from the 1980s onward, carried by influential work such as that of Pfeffer and Salancik (1978) and DiMaggio and Powell (1983). Economics became increasingly important, starting in 1975 with the work of Williamson (1975). Its importance continued to rise, peaking between 1991 and 1993 in particular because of the work of Amos Tversky and Daniel Kahneman birthing behavioral economics. Statistical and methodological work has been steadily influential from the 1940s to the present. Other perspectives such as mathematics, biology, and anthropology have had only a small influence on the field.

Most influential articles

In addition to the major themes and disciplinary perspective of the field, and their evolution, it is useful to study the most influential works too, to understand which ideas gripped the field. Indeed, these works were often seminal to new research programs. To understand in greater detail the articles with the most influence on the field, I tallied the number of times each of the 225 unique articles were in the top five most cited articles of one of the clusters, across the 16 time periods. The top 20 most important articles are displayed in Table 17. This analysis provides a more accurate assessment of the most influential article than simply looking at the number of times an article is cited. Indeed, an article can have an outsized influence on a smaller research area, while garnering fewer citations than less influential articles in more highly active research areas. Furthermore, using raw citations as a measure of influence favors older articles. As a general tendency, we can see that most of the highly influential management articles were published between 1970 and 1990.

[Insert Table 17 about here]

Discussion

Trends and research programs

As discussed in the introduction, the management literature has been increasing non-linearly (see Figure 5). This fact is explained by at least two factors: (a) by a growth of the scientific community, or (b) by a growth of interest in the field. Although this research does not provide evidence for (a), I argue that the multidisciplinary nature of management provides evidence for (b). In total, I've identified 8 different disciplinary perspectives that were used for research.

From 1940 to 1970

Figure 6 shows that, until the mid-to-late 80s, psychology was the dominant perspective employed to study management phenomena. This result is unsurprising because most of the earliest scientific journals that published management research were psychologically oriented. Indeed, before 1954, the four outlets for management were the Journal of Applied Psychology, Harvard Business Review, Human Relations, and Personnel Psychology. In 1954, Management Science entered print, followed by Administrative Science Quarterly in 1956, and Academy Management Journal in 1958. From 1958 onwards, most new journals seem to become more management-focused (see Table 12). This gradual shift away from psychology as the dominant perspective, towards management, may reflect the fact that management, as its own discipline, became truly legitimate in the 1950-1960 period.

In terms of the topics that were studied during the 1940 to 1960 period, most research can be classified into two categories: either researchers were interested in the correlates, psychological or otherwise, that are associated with worker productivity, or they were interested in creating measurement instruments and diagnostic tools to better increase person-job fit. For instance, some of the work in clusters 1 and 9 of this period fit into the first category because it deals with overcoming resistance to change or identifying the differences in leadership between high- and low-productivity teams. Other clusters, such as clusters 3, 5, and 7 fit into the second category. Research interests shift away from measurement instruments in the 1961-1970 period. While the work of Fleishman in cluster 9 is mostly focused on the effect of leaders on worker productivity, most research now focuses on the inner working of the mind and, with the work of Herzberg and Maslow in particular, on motivation.

From 1970 to 1999

Beginning in the 1971-1980 period, we can see the appearance of a new research stream: Organizational Theory. This research area, broadly speaking, takes the organization as the unit of analysis and examines its behavior as a function of internal or external factors. The emergence of

Organization Theory as a research program marks the emergence of a new dimension of management research: research taking the individual or the organization as the unit of analysis. This finding is corroborated by others who make distinctions between micro and macro research (Agarwal & Hoetker, 2007; Schminke & Mitchell, 2003). This axis can be seen visually in the co-citation network visualization of the 1971-1980 period in Appendix 3. On the right side of the network, we can see the Organizational theory cluster emerging out of the "New Management" cluster.

Between 1980 and 1990, many considered management departments to be the repositories of multidisciplinary research (Bennis & O'Toole, 2005; Corley & Gioia, 2011) where several disciplinary perspectives were used, and valued, to investigate the same phenomenon. For instance, the Strategic Management cluster from the 1981-1985 period (see Appendix 3) contains at once management, sociological, and psychological perspectives. Porter (1980) and Chandler Jr (1962) are both interested in the efficient and effective administration of an organization. Both Thompson (1967) and Child (1972) study the behavior and structure of the organization as a function of its context and March and Simon (1958) give useful information on how to structure the organization to facilitate information processing, which has important implications for strategic management. Despite multidisciplinarity, the rift between the individual or the organization as a unit of analysis continues to increase, as seen in the co-citation network for the 1981-1985 period, and even more so in the 1986-1990 period (see Appendix 3). On the left side of the network visualization, we have research taking the organization as the unit of analysis, and on the right, research taking the individual as the unit of analysis with little communication between the two.

Starting in 1991, we see the emergence of a new research area. This area, which Hambrick and Mason (1984) call the "upper echelons perspective" focuses on the relationship between the characteristics of top management and organizational outcomes. Interestingly, the "upper echelons" literature is situation closer to research areas focusing on the organization (see 1991-1993 period in Appendix 3), perhaps due to the upper echelons perspective focusing largely on organizational outcomes rather than the characteristics of upper management.

Besides this emergence, research interests remain stable between 1991 and 1999 and there is little change in the shapes of the co-citation networks. Despite this stability, the rift between individuals and organizations as the unit of analysis continues to solidify. Furthermore, we begin seeing a reduction in the multidisciplinarity of management research, a trend that was rising since the 1970s. Indeed, Figure 6 shows that the influence of psychology, sociology, and economics on management research has been trending downwards since about 1991. This downward trend may be explained by borrowing from the work of Gort and Klepper (1982) and others, on the diffusion of innovation. They explain how, in the beginning, a nascent industry has very little industry-specific knowledge, and thus has a malleable institutional environment that needs legitimacy (Aldrich & Fiol, 1994; Hannan & Freeman, 1977). To gain legitimacy, a nascent industry can borrow knowledge and practices from related industries. As the nascent industry acquires legitimacy and matures, its stock of industry-specific knowledge, rules, and routines increases (Gort & Klepper, 1982), and the value of knowledge from outside the industry's boundaries become less relevant. Analogously, in its early years, management turned to related disciplines not only for their theories but also for their methods. As the field matured and developed its own theories and methods, the need for borrowing decreased (Agarwal & Hoetker, 2007).

From 2000 to 2022

Between the years 2000 and 2002, a new and important research area emerged: Strategic Human Resource Management, pioneered by the work of Huselid (1995). From the 2000s onwards, Strategic Human Resource Management, along with Organizational Psychology, Organizational Theory, and Strategic Management become staples of Management research, and interest in these areas continues in 2022. Sometimes, in addition to the four staples, a new research area appears, such as Embeddedness and Organization Networks between 2003 and 2005, but is quickly subsumed into another cluster. Inspecting the co-citation networks in Appendix 3, from the 2000s to 2022 reveals an interesting insight: the Strategic Human Resource Management literature is situated between research focused on the individual, and research focused on the organization. This may indicate that Strategic Human Resource Management bridges this divide by tying individual psychology to organizational performance.

Another interesting trend, which truly set in the 2012-2014 period but began in the 2000s, is a gradual disconnect between Organization theory and Organizational psychology. This disconnect marks the emergence of a new dimension: the qualitative-quantitative dimension. Indeed, from 2012 onward, quantitative and qualitative methods become the most cited articles in the Organizational psychology and Organization theory literature respectively. In addition, this disconnect is seen in the entire clusters, not only in the top-cited articles. Of the 824 articles contained in the Organizational psychology cluster, a majority are quantitative, whereas, of the 432 articles contained in the Organizational theory cluster, a majority are qualitative.

The trends identified above continue today. If we look at the 2018-2020 and 2021-2022 co-citation networks, we can see the four staple research areas, active since the 2000s, the Organization-Individual axis, as well as the Qualitative-Quantitative axis. In addition, we can see that "hybrid" perspectives such as the Upper Echelons perspective in the 2018-2020 period, Research methods in the 2021-2022 period, and Strategic Human Resource Management in both periods bridge the Organization-Individual divide and even the Qualitative-Quantitative divide to a limited degree.

Most influential authors and ideas

Description of the 20 most influential articles

Table 17 shows the 20 most influential articles in my selected sample. These articles are influential because they are one of the top 5 most cited articles of one cluster. To better understand the ideas that had an outsized influence on the field, I've summarized the core ideas of these works in Table 18. We can see that most of these selected works are dedicated to understanding the behavior of organizations. These works view the organization in different ways; as dependent on its environment for survival (Pfeffer & Salancik, 1978), as a bundle of resources (Barney, 1991), or as a nexus of contracts (Jensen & Meckling, 1976). These works seek to understand the behavior, capabilities, and structure of the organization. The remainder is split into two areas. One part focuses on the psychology of individuals (Hofstede, 1980; Vroom, 1964), while another is about qualitative and quantitative methods (Aiken, West, & Reno, 1991; Baron & Kenny, 1986; Eisenhardt, 1989; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003).

[Insert Table 18 about here]

As discussed in the introduction, management is traversing a crisis. Many of our theories are of dubious validity (Aguinis & Vandenberg, 2014; Antonakis, 2017; Ferris et al., 2012; Tourish, 2020; Van de Ven, 1989), and most have never been empirically tested (Edwards, 2010; Edwards et al., 2014; Kacmar & Whitfield, 2000), and those that have are not assembled into programmatic theories (Aguinis & Cronin, 2022; Cronin et al., 2021). For example, building on an earlier review (Miner, 1984), Miner (2003) shows that out of 73 reviewed management theories, only 25 were rated as high in scientific validity. On the empirical side, our literature is plagued by endogeneity, whereby about 80 to 90% of articles contain at least one threat to validity (see Table 11), which leads to biased, causally uninterpretable results (Antonakis et al., 2021; Antonakis et al., 2010; Fischer et al., 2017; Hamilton & Nickerson, 2003). Unfortunately, these issues span the entire field, and the 20 most important articles are no exception. The fact that the most influential theories also contain theoretical and empirical problems highlights the importance of course correction. In what follows, I will provide a critique of two of the 20 most important articles, highlighting common problems in theoretical and empirical research.

Critique of two exemplars: theory building and theory testing

In this section, I critique two articles that are both exemplars of contributions in theory building and theory testing and contain common theoretical and empirical problems found in the management literature. I begin with a critique of the Resource-Based View, introduced by Barney (1991), showing that his theoretical arguments are, in part, tautological. Then, I critique the work of Huselid (1995), which was seminal to the Strategic Human Resource Management literature, arguing that the statistical models presented are uninterpretable because of endogeneity.

Critique of Barney (1991)

In many ways, Barney (1991) in his expository article introduced the resource-based view (RBV) of the firm. This view has grown in popularity, and his seminal article garners now more than 85,000 citations on google scholar and is one of the most cited macro, or organizational level theories to date (Colquitt & Zapata-Phelan, 2007). In what follows I will give a summary of the RBV, and offer some critiques of it as a theoretical system. For a detailed critique of the RBV see the work of Priem and Butler (2001), which inspired this summary.

The RBV rests on two fundamental axiomatic assumptions: (1) resources are distributed heterogeneously across firms, and (2) these productive resources cannot be transferred from firm to firm without cost (Barney, 1991). Given these axioms, Barney (1991) offers two theoretical propositions:

1. Resources that are both rare (i.e., held by few firms) and valuable (i.e., contribute to the efficiency and effectiveness of the firm) can produce a competitive advantage.
2. When rare and valuable resources are also not imitable, not substitutable, and not transferrable those resources can produce a *sustainable* competitive advantage.

I will now evaluate these two propositions considering the three necessary criteria for a scientific theory which are: (1) Generalized conditionals (i.e., if/then statements), (2) Empirical content (i.e., statements which require empirical testing and are not true by definition) and (3) nomic necessity (i.e., that two variables are causally related). First, the RBV clearly contains generalized conditionals. For example, this view states that *if* a firm resource or attribute is rare and valuable, *then* that resource or attribute can give the firm a competitive advantage. Moreover, if such a resource is nonsubstitutable and hard to imitate, then it can provide the firm with a sustainable competitive advantage. Second, to examine the empirical content of a statement, to determine if it is analytical or synthetic, it is often useful to replace the terms in that statement with their definition. Let us examine the first theoretical proposition: resources that are both rare and valuable can produce a competitive advantage (Barney, 1991, p. 107). He defines firm resources as: "firm attributes that may enable firms to conceive of and implement value-creating strategies" (Barney, 1991, p. 101). Furthermore, he defines firm resources as *valuable* when: "they enable a firm to conceive of or implement strategies that improve its efficiency and effectiveness" (Barney, 1991, p. 106). Finally, he says that a firm has a competitive advantage when: "it is implementing a value-creating strategy not simultaneously being implemented by any current or potential competitors." (Barney, 1991, p. 102). Thus, substituting these definitions in the first theoretical proposition yields: "*Firm attributes that may enable firms to conceive of and implement value-creating strategies* that are both rare and *that enable a firm to conceive of or implement strategies that improve its efficiency and effectiveness* can produce a *value-creating strategy not simultaneously being implemented by any current or potential competitors*." Substituting the definitions of firm ressources, valuable ressources and competitive advantage into the first proposition reveals that the RBV is an analytical statement, it is true by definition because it is tautological. Third, the basic statements of the RBV cannot be examined for nomic necessity, because the statement is tautological. This simple analysis reveals that the Resource-Based View, as exposed in (Barney, 1991) cannot be a fruitful theoretical perspective.

Critique of Huselid (1995)

The work of Huselid (1995) is considered groundbreaking by many because it showed a link between a set of HR practices termed high-performance work systems (HPWS) and turnover, productivity, and corporate profits (Gerhart, Wright, & McMahan, 2000; Kaufman, 2010; Paauwe, 2009; Wright, Gardner, Moynihan, & Allen, 2005). His work has also been the recipient of considerable amounts of criticism on topics such as measurement error (Gerhart et al., 2000), conceptual soundness (Kaufman, 2015), and the direction of causality in his model (Wright et al., 2005). The premise of his study is that systems of *High-Performance Work Practices* (HPWP), which are bundles of individual Human Resources practices, can have important effects on employee turnover, productivity, and overall corporate financial performance (Huselid, 1995). Furthermore, the degree of complementarity between the individual human resource practices that make up the overall HPWP bundle, as well as the alignment between the HPWP bundle and the overall corporate strategy moderates these relationships. The basic model, omitting control and moderating variables is presented in Figure 7. Tautologies, similar to those presented above, could be present because the strategic human resource paradigm is largely based on the RBV (Allen & Wright, 2007; Wright, Dunford, & Snell, 2001). However, in what follows, I will focus solely on methodology.

[Insert Figure 7 about here]

The main issue in Huselid (1995) comes from the fact that HPWPs are endogenous. Looking at Figure 7, it is easy to imagine a multitude of omitted variables (see Table 11) that may simultaneously cause both HPWP (the independent variable) and turnover, productivity, and corporate financial performance (the dependent variable). For example, high-quality corporate leadership, because of their intelligence, experience, or skill, may choose to implement bundles of HPWP and cause a low turnover and financial performance simultaneously. Similarly, corporate culture, if it is focused on efficiency, may be the cause of both HPWP and employee productivity. When omitted variables are causes of both the independent and dependent variables, the estimated regression coefficients become difficult to interpret: they may be upwardly or downwardly biased, and may even be of a different sign (Antonakis et al., 2010; Ketokivi & McIntosh, 2017). Detecting endogeneity is possible with a Hausman test (Hausman, 1978). If endogeneity is present, it can be corrected for using instrumental variables estimation (Wooldridge, 2002 (see chapter 5); 2013 (see chapter 15)). Critically, both detecting and correcting for endogeneity requires instrumental variables. Instrumental variables are variables that do not correlate with the omitted cause, that is they are uncorrelated with the structural error term in a model, but are correlated with the endogenous independent variable (Antonakis et al., 2010; Ketokivi & McIntosh, 2017). Huselid (1995, p. 666) makes a case that endogeneity due to simultaneity is not a concern. He bases this case on the non-significant results of a Hausman test. However, he does not present nor discuss exogenous instrumental variables, thus it is difficult to see how a proper Hausman test could have been conducted.

In all, I believe the critiques I've presented help highlight common problems in management theorizing and empirical work. It is only by recognizing the current problems in our literature that we can move the field towards more meaningful theories, sound empirical work, and ultimately consequential, ethical, and accurate policy recommendations.

Limitations & Future Directions

Like all work, mine is not without its limitations. Whereas co-citation network analysis excels in identifying core research programs, which was the aim of this research, it is ill-suited to identify smaller, niche research programs. Other methods, such as bibliographic coupling are better suited for this purpose. Because core areas were once niche, future research could focus on niche research programs that are active today, to detect the trends of tomorrow.

Because some elements of a co-citation network analysis are closer to art than science (Zupic & Čater, 2015), my analysis contains an element of subjectivity. For example, the widths of the periods were selected so that each period would accurately match the speed at which research advances, while also avoiding redundant analyses. Other ways of splitting periods could have been chosen. Furthermore, I chose to adjust the cutoff for inclusion into the network such as to include the top 1% of the most cited articles into this analysis. Other cutoffs could have been chosen such as 2% or 5%. Nevertheless, when conducting preliminary analysis to determine the final cutoff value, a cutoff of 2% or 5% did not change the results substantially. Unfortunately, computational limitations also limited the number of articles that could be included in the analysis, particularly in the 2018-2020 period.

Noteworthy too, as discussed above, is that citations counts are noisy indicators of an articles' influence and high citation counts do not necessarily translate into consequential work. Indeed, authors cite work for various reasons and citations may be "boiler-plate" and meant to indicate priviness to specific literatures. Therefore, operationalizing the influence of an article by its citation count may lead to including highly cited but unimportant work, while excluding low citation count but highly influential articles from our sample.

An element of subjectivity also entered my analysis when identifying the theme of each cluster. Indeed, co-citation analysis simply groups items according to citation patterns but does not indicate why these items were co-cited. To identify each cluster, I analyzed the title and abstract of the five most cited items in each cluster. These themes were therefore extracted based on my understanding of the factors that united these items into a cluster. Although this method provides objectivity, future work could use natural language processing methods to identify the most frequent words in the abstracts and titles of all items included in a cluster. Finally, the construction of Figure 6 was based on a classification of items into a disciplinary category. Whereas most items clearly employed a disciplinary perspective, others were multidisciplinary in nature. In those cases, my classification was based on the most dominant perspective, which may not fully capture the breadth of some work.

Conclusion

In conclusion, the field of management has greatly grown and evolved over the studied period. It largely began as an outgrowth of psychology, but rapidly became its own disciplinary perspective. Beginning in the 1980s, other disciplinary perspectives such as economics and sociology began to be used to investigate management phenomena leading to a new dimension characterizing the field: research focused on the individual or the organization. Today, the field of management is largely comprised of four core mostly self-contained research programs: Organizational Theory, Strategic Management, Organizational Psychology, and Strategic Human Resource Management. Furthermore, there appears to be a new dimension of research dependent on the choice of qualitative or quantitative methods.

As noted above, the management literature is traversing a crisis. Our theories are of dubious validity (Aguinis & Vandenberg, 2014; Antonakis, 2017; Ferris et al., 2012; Miner, 1984, 2003; Tourish, 2020; Van de Ven, 1989), and most have never been empirically tested (Edwards, 2010; Edwards et al., 2014; Kacmar & Whitfield, 2000). Furthermore, our empirical literature is plagued by endogeneity, whereby about 80 to 90% of articles contain at least one threat to validity (see Table 11), which leads to biased results, hindering causal interpretability (Antonakis et al., 2021; Antonakis et al., 2010; Fischer et al., 2017; Hamilton & Nickerson, 2003). We therefore call on scholar to take inspiration from more mature, fundamental disciplines such as economics, psychology or biology which build upon a set of core, foundational theories. Indeed, as a science our ultimate goal is a unified management literature composed of theoretically and empirically sound programmatic theories (Aguinis & Cronin, 2022; Cronin et al., 2021). The only way to move our field towards this goal is to acknowledge the theoretical and empirical problems our literature is currently facing, some of which have been discussed above, and to address them.

Main Text References

Agarwal, R., & Hoetker, G. (2007). A Faustian bargain? The growth of management and its relationship with related disciplines. *Academy of management journal, 50*(6), 1304-1322. doi:10.5465/AMJ.2007.28165901

Aguinis, H., & Cronin, M. A. (2022). It's the Theory, Stupid. *Organizational Psychology Review*, 20413866221080629.

Aguinis, H., & Vandenberg, R. J. (2014). An Ounce of Prevention Is Worth a Pound of Cure: Improving Research Quality Before Data Collection. *Annual Review of Organizational Psychology and Organizational Behavior, Vol 1, 1*, 569-595. doi:10.1146/annurev-orgpsych-031413-091231

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*: sage.

Aldrich, H. E., & Fiol, C. M. (1994). Fools rush in? The institutional context of industry creation. *Academy of Management review, 19*(4), 645-670.

Allen, M., & Wright, P. (2007). Strategic Management and HRM. In P. Boxall, J. Purcell, & P. Wright (Eds.), *The Oxford Handbook of Human Resource Management* (pp. 88-107): Oxford University Press.

Antonakis, J. (2017). On doing better science: From thrill of discovery to policy implications. *Leadership Quarterly, 28*(1), 5-21. doi:10.1016/j.leaqua.2017.01.006

Antonakis, J., Bastardoz, N., Liu, Y., & Schriesheim, C. A. (2014). What makes articles highly cited? *The Leadership Quarterly, 25*(1), 152-179.

Antonakis, J., Bastardoz, N., & Rönkkö, M. (2021). On ignoring the random effects assumption in multilevel models: Review, critique, and recommendations. *Organizational Research Methods, 24*(2), 443-483.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *Leadership Quarterly, 21*(6), 1086-1120. doi:10.1016/j.leaqua.2010.10.010

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management review, 14*(4), 496-515.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99-120.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173.

Bennis, W. G., & O'Toole, J. (2005). How business schools have lost their way. *Harvard business review, 83*(5), 96-104.

Chandler Jr, A. D. (1962). *Strategy and structure: chapters in the history of the industrial enterprise*. Oxford, England: M.I.T. Press.

Child, J. (1972). Organizational structure, environment and performance: The role of strategic choice. *sociology, 6*(1), 1-22.

Colquitt, J. A., & Zapata-Phelan, C. P. (2007). Trends in theory building and theory testing: A five-decade study of the Academy of Management Journal. *Academy of management journal, 50*(6), 1281-1303.

Corley, K. G., & Gioia, D. A. (2011). Building theory about theory building: what constitutes a theoretical contribution? *Academy of Management review, 36*(1), 12-32.

Cronin, M. A., Stouten, J., & van Knippenberg, D. (2021). The theory crisis in management research: Solving the right problem. *Academy of Management review, 46*(4), 667-683.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American sociological review*, 147-160.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research, 133*, 285-296.

Edwards, J. R. (2010). Reconsidering theoretical progress in organizational and management research. *Organizational Research Methods, 13*(4), 615-619.

Edwards, J. R., Berry, J., & Kay, V. S. (2014). *Bridging the great divide between theoretical and empirical management research.* Paper presented at the Academy of Management Proceedings.
Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management review, 14*(4), 532-550.
Ferris, G. R., Hochwarter, W. A., & Buckley, M. R. (2012). Theory in the organizational sciences: How will we know it when we see it? *Organizational Psychology Review, 2*(1), 94-106.
Fischer, T., Dietz, J., & Antonakis, J. (2017). Leadership process models: A review and synthesis. *Journal of Management, 43*(6), 1726-1753.
Gerhart, B., Wright, P. M., & McMahan, G. C. (2000). Measurement error in research on the human resources and firm performance relationship: Further evidence and analysis. *Personnel Psychology, 53*(4), 855-872.
Gordon, R. A., & Howell, J. E. (1959). *Higher Education for Business*: New York: Columbia University Press.
Gort, M., & Klepper, S. (1982). Time paths in the diffusion of product innovations. *The economic journal, 92*(367), 630-653.
Hambrick, D. C., & Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management review, 9*(2), 193-206.
Hamilton, B. H., & Nickerson, J. A. (2003). Correcting for endogeneity in strategic management research. *Strategic organization, 1*(1), 51-78.
Hannan, M. T., & Freeman, J. (1977). The population ecology of organizations. *American journal of sociology, 82*(5), 929-964.
Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 1251-1271.
Hofstede, G. (1980). Culture's consequences: International differences in work-related values. In: sage.
Hunt, S. (1991). Modern marketing theory: Critical issues in the philosophy of marketing science.
Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of management journal, 38*(3), 635-672.
Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial economics, 3*(4), 305-360.
Kacmar, K. M., & Whitfield, J. M. (2000). An additional rating method for journal articles in the field of management. *Organizational Research Methods, 3*(4), 392-406.
Kaufman, B. E. (2010). SHRM theory in the post-Huselid Era: why it is fundamentally misspecified. *Industrial Relations: A Journal of Economy and Society, 49*(2), 286-313.
Kaufman, B. E. (2015). Market competition, HRM, and firm performance: The conventional paradigm critiqued and reformulated. *Human Resource Management Review, 25*(1), 107-125.
Ketokivi, M., & McIntosh, C. N. (2017). Addressing the endogeneity dilemma in operations management research: Theoretical, empirical, and pragmatic considerations. *Journal of Operations Management, 52*, 1-14.
Kuhn, T. S. (1970). *The structure of scientific revolutions* (Vol. 111): Chicago University of Chicago Press.
Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. *Journal of Operations Management, 64*, 19-40.
Mantere, S., & Ketokivi, M. (2013). Reasoning in organization science. *Academy of Management review, 38*(1), 70-89.
March, J. G., & Simon, H. A. (1958). *Organizations*: John wiley & sons.
McCain, K. W. (1990). Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science (1986-1998), 41*(6), 433.
McKelvey, B. (1997). Perspective—Quasi-natural organization science. *Organization science, 8*(4), 351-380.
Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological inquiry, 1*(2), 108-141.
Miner, J. B. (1984). The validity and usefulness of theories in an emerging organizational science. *Academy of Management review, 9*(2), 296-306.

Miner, J. B. (2003). The rated importance, scientific validity, and practical usefulness of organizational behavior theories: A quantitative review. *Academy of Management Learning & Education, 2*(3), 250-268.
Mintzberg, H. (2005). Developing theory about the development of theory in KG Smith & MA Hitt (Eds) Great minds in management: The process of theory development (pp. 355-372). In: Oxford: Oxford University Press.
Paauwe, J. (2009). HRM and performance: Achievements, methodological issues and prospects. *Journal of Management studies, 46*(1), 129-142.
Pasadeos, Y., Phelps, J., & Kim, B.-H. (1998). Disciplinary impact of advertising scholars: Temporal comparisons of influential authors, works and research networks. *Journal of Advertising, 27*(4), 53-70.
Peirce, C. (1878). Deduction, induction, and hypothesis. . *Popular Science Monthly, 13*, 470-482.
Pfeffer, J., & Salancik, G. R. (1978). *The external control of organizations: A resource dependence perspective*: Stanford University Press.
Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879.
Porter, M. E. (1980). Competitive strategy. *Measuring business excellence*.
Priem, R. L., & Butler, J. E. (2001). Is the resource-based “view” a useful perspective for strategic management research? *Academy of Management review, 26*(1), 22-40.
Rudner, R. S. (1966). Philosophy of Social Science.
Sankey, H. (2013). Scientific method. In *The Routledge companion to philosophy of science* (pp. 312-322): Routledge.
Schminke, M., & Mitchell, M. (2003). From the Editors. In the Beginning…. In (Vol. 46, pp. 279-282): Academy of Management Briarcliff Manor, NY 10510.
Silk, L. S. (1964). The goals of business education. *The Journal of Risk and Insurance, 31*(3), 421-428.
Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265-269.
Thompson, J. D. (1967). *Organizations in action: Social science bases of administrative theory*: Transaction publishers.
Tourish, D. (2020). The triumph of nonsense in management studies. *Academy of Management Learning & Education, 19*(1), 99-109.
Van de Ven, A. H. (1989). Nothing is quite so practical as a good theory. *Academy of Management review, 14*(4), 486-489.
Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *scientometrics, 84*(2), 523-538.
Van Eck, N. J., & Waltman, L. (2007). Bibliometric mapping of the computational intelligence field. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 15*(05), 625-645.
Van Eck, N. J., Waltman, L., Dekker, R., & Van Den Berg, J. (2010). A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology, 61*(12), 2405-2416.
van Eck, N. J., Waltman, L., van den Berg, J., & Kaymak, U. (2006). Visualizing the computational intelligence field [Application Notes]. *IEEE Computational Intelligence Magazine, 1*(4), 6-10.
Vroom, V. H. (1964). Work and motivation.
Waltman, L., Van Eck, N. J., & Noyons, E. C. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of informetrics, 4*(4), 629-635.
Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management review, 14*(4), 490-495.
Williamson, O. E. (1975). Markets and hierarchies: analysis and antitrust implications: a study in the economics of internal organization. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*.
Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*: MIT Press.
Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.): South-Western College.

Wright, P. M., Dunford, B. B., & Snell, S. A. (2001). Human resources and the resource based view of the firm. *Journal of Management*, *27*(6), 701-721.

Wright, P. M., Gardner, T. M., Moynihan, L. M., & Allen, M. R. (2005). The relationship between HR practices and firm performance: Examining causal order. *Personnel Psychology*, *58*(2), 409-446.

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, *18*(3), 429-472.

Figures

Figure 5

The number of articles published in 3-year periods, in the top 50th percentile core-management journals from 1940 to 2020, as well as the number of cited references per article.

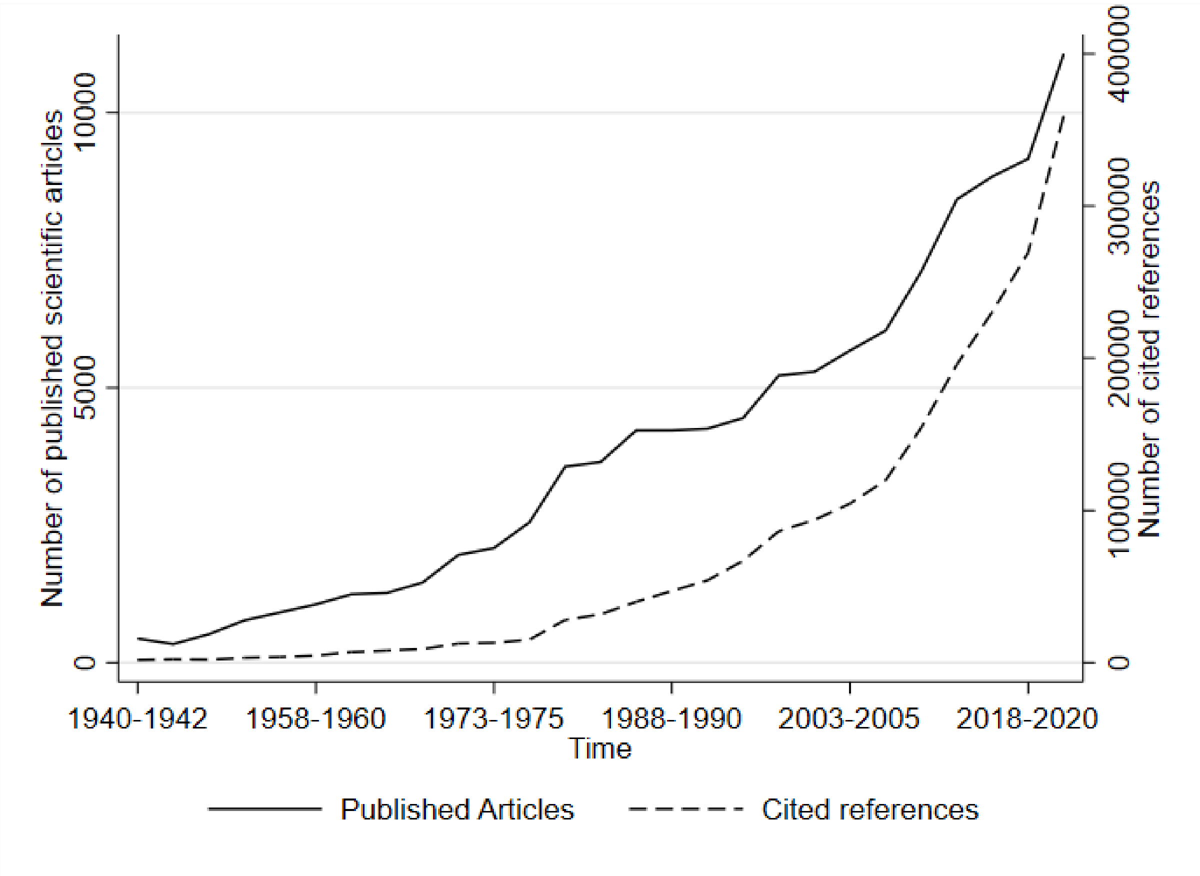
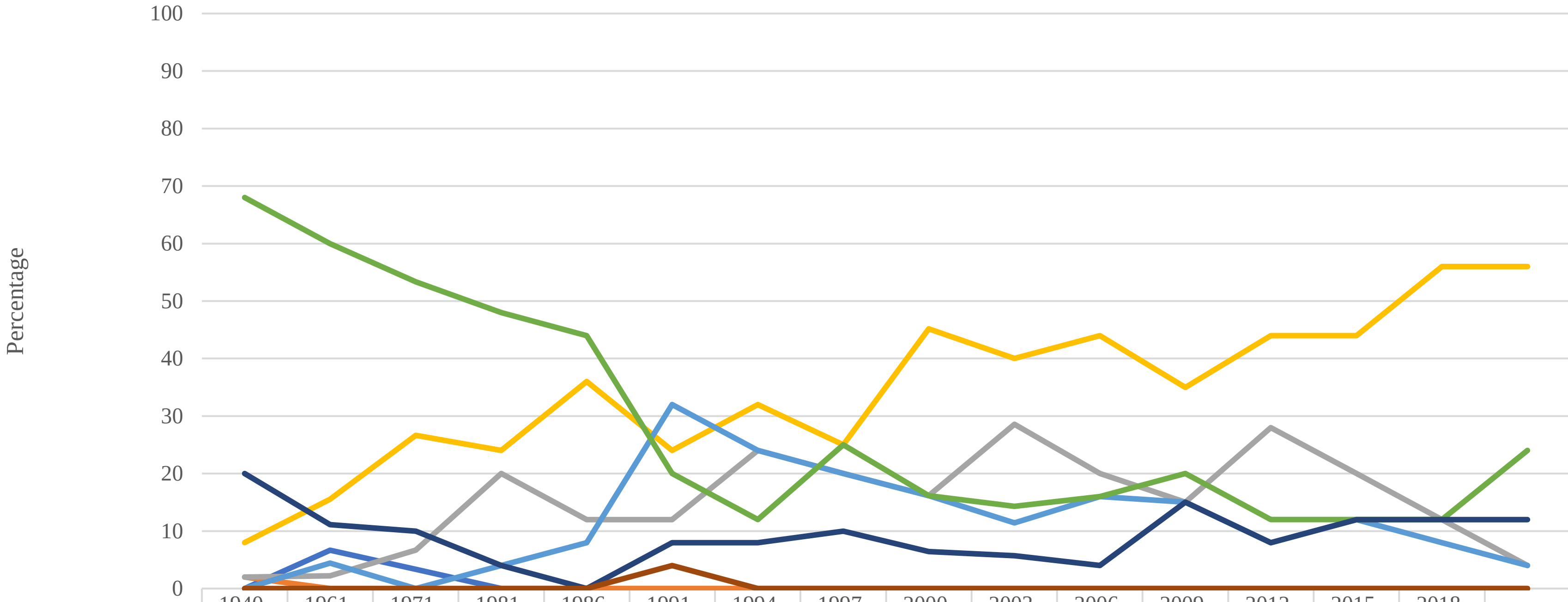


Figure 6

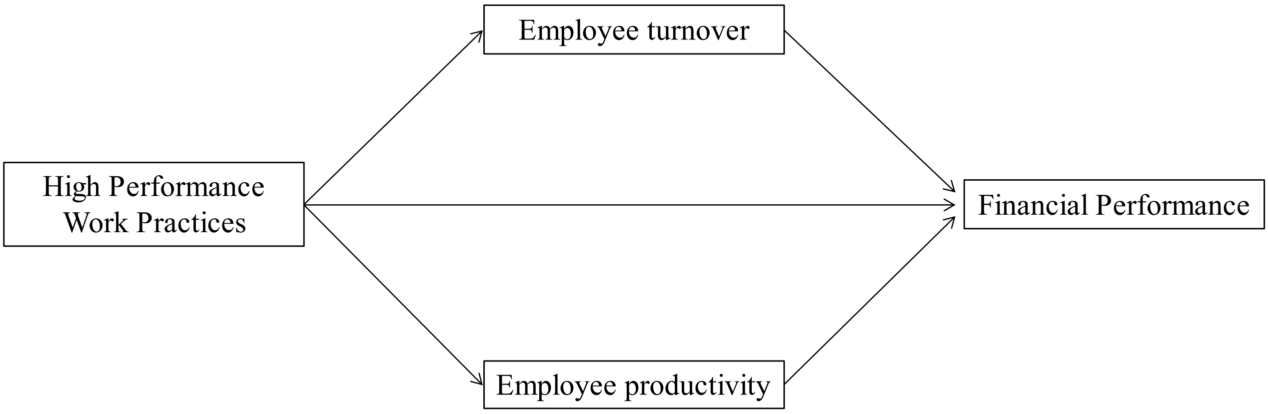
Contribution of different disciplinary perspectives identified to the management literature, from 1940 to 2022



	1940-1960	1961-1970	1971-1980	1981-1985	1986-1990	1991-1993	1994-1996	1997-1999	2000-2002	2003-2005	2006-2008	2009-2011	2012-2014	2015-2017	2018-2020	2021-P
Mathematics	0	7	3	0	0	0	0	0	0	0	0	0	0	0	0	0
Anthropology	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sociology	2	2	7	20	12	12	24	20	16	29	20	15	28	20	12	4
Management	8	16	27	24	36	24	32	25	45	40	44	35	44	44	56	56
Economics	0	4	0	4	8	32	24	20	16	11	16	15	8	12	8	4
Psychology	68	60	53	48	44	20	12	25	16	14	16	20	12	12	12	24
Statistics	20	11	10	4	0	8	8	10	6	6	4	15	8	12	12	12
Biology	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0

Figure 7

The core model presented in Huselid (1995), omitting control variables and measures of internal and external fit.



Tables

Table 11

The 7 threats to the validity of quantitative research, adapted from (Antonakis, Bendahan, Jacquart, & Lalive, 2010)

Threat to validity	Explanation
1. Omitted variables	Omitting a regressor, a fixed-effect in a multilevel model, using random effects without justification, or if the independent variable is not exogenous
2. Omitted selection	Comparing the groups to which observations have not been randomly assigned, comparing groups to which assignment was endogenous, comparing groups with self-selection
3. Simultaneity	The independent and dependent variables simultaneously cause each other, reverse causality
4. Measurement error	Failing to model and correct for variables measured with error
5. Common-method variance	Both independent and dependent variables are gathered from the same source
6. Inconsistent inference	Failing to use robust standard errors in the presence of heteroskedasticity, not using cluster-robust standard errors in panel data[a]
7. Model misspecification	Failing to correlate disturbance in endogenous mediation models, using a full information estimator without comparing estimates to a limited information estimator

Note: (a) failing to use robust or cluster-robust standard errors does not affect the consistency of the estimate, but of the standard errors causing the corresponding *p*-value to be over or understated.

Table 12

Top 50th percentile core-management journals included in our search for management articles, with 2020 journal impact factor, sorted by year of first publication,

Journal name	2020 JIF	Year of 1st publication
Journal Of Applied Psychology	7.43	1917
Harvard Business Review	6.87	1922
Human Relations	5.73	1947
Personnel Psychology	7.07	1948
Management Science	4.88	1954
Administrative Science Quarterly	11.11	1956
Academy Of Management Journal	10.19	1958
California Management Review	8.84	1958
Journal Of Management Studies	7.39	1961
Human Resource Management	5.08	1962
Journal Of Small Business Management	4.54	1963
Management Decision	4.96	1967
Long Range Planning	8.80	1968
Journal Of International Business Studies	11.38	1970
Research Policy	8.11	1972
Journal Of Management	11.79	1975
Academy Of Management Review	12.64	1976
Strategic Management Journal	8.64	1980
Organization Studies	6.31	1980
European Management Journal	5.08	1982
International Journal Of Project Management	7.17	1983
Organizational Behavior And Human Decision Processes	4.94	1985
Academy Of Management Perspectives	7.85	1987
Journal Of Organizational Behavior	8.17	1988
Small Business Economics	8.16	1989
Leadership Quarterly	10.52	1990
British Journal Of Management	6.57	1990
International Journal Of Human Resource Management	5.55	1990
Organization Science	5.00	1990
Human Resource Management Review	7.44	1991
Human Resource Management Journal	5.04	1991
European Journal Of Work And Organizational Psychology	3.97	1991
Business Strategy And The Environment	10.30	1992
Journal Of Occupational And Organizational Psychology	4.56	1992
Corporate Social Responsibility And Environmental Management	8.74	1994
Organization	5.12	1994
Journal Of Knowledge Management	8.18	1996
International Journal Of Management Reviews	13.42	1999
Human Resource Development Review	4.74	2002
Strategic Organization	5.41	2003
Academy Of Management Annals	16.44	2007
Strategic Entrepreneurship Journal	9.29	2007
Annual Review Of Organizational Psychology And Organizational Behavior	18.33	2014
Journal Of Innovation & Knowledge	9.27	2016

Table 13

Time ranges selected for analysis, number of scientific articles and cited references retrieved in each period, minimum citations of a reference for inclusion in the co-citation network, and number of included references in the co-citation network.

Time range	Number of scientific articles retrieved	Number of cited references	Minimum citations for inclusion in the network	Number of included references in the network
1940-1960	5,270	21,866	6	218
1961-1970	5,274	28,879	7	322
1971-1980	9,995	64,710	11	636
1981-1985	6,920	61,784	11	668
1986-1990	7,053	79,907	14	833
1991-1993	4,448	67,007	14	630
1994-1996	5,219	86,387	14	898
1997-1999	5,287	93,818	15	923
2000-2002	5,673	104,490	16	1,079
2003-2005	6,037	119,853	17	1,194
2006-2008	7,114	154,416	20	1,492
2009-2011	8,423	195,846	22	1,976
2012-2014	8,841	230,582	22	2,425
2015-2017	9,160	269,455	23	2,733
2018-2020	11,071	359,539	24	2,499
2021-2022	4,818	215,608	14	2,042

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *Leadership Quarterly*, *21*(6), 1086-1120. doi:10.1016/j.leaqua.2010.10.010

Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of management journal*, *38*(3), 635-672.

Table 14

Number of clusters, major theme, and percentage of references dedicated to each theme in the periods from 1940 to 1999

1940-1960	%	1961-1970	%	1971-1980	%	1981-1985	%	1986-1990	%	1991-1993	%	1994-1996	%	1997-1999	%
Management	22.5	Research Methods	27.0	Organizational theory	28.8	Motivation and job satisfaction	30.1	Motivation and Job satisfaction	40.8	Organizational psychology	38.7	Organizational psychology	36.7	Organizational psychology	37.1
Research Methods	15.6	"New management", organizational sociology	26.7	Motivation and Job satisfaction	27.8	Strategic management	23.2	Organizational theory	26.5	Organizational theory	30.0	Organizational structure	23.7	Organizational structure, resource dependence	25.4
Occupational counselling	15.6	Work motivation	14.9	Research Methods	17.5	Organizational theory	20.4	Strategic management	19.4	Organizational strategy and structure	21.9	Strategic Management	15.9	Organizational theory	19.6
Job satisfaction	8.7	Group psychology	11.2	"New management"	15.6	Research Methods	14.8	Organizational structure	12.7	Heuristics, Biases and Decision making	4.9	Organizational Theory	12.2	Strategic management/Upper echelons perspectives	18.0
Vocational interests	8.7	Linear programming, optimization	5.9	Human performance (in the organization)	7.9	Organizational design	11.5	The "trapped administrator"	0.5	Upper echelons perspectives	4.4	Upper echelons perspectives	11.4		
Job characteristics	6.9	Need satisfaction	5.0	Managerial decision making	2.5										
Multiphasic personality inventory	6.0	Job interviews	3.4												
Group psychology	5.5	Factors influencing productivity	3.4												
Leadership	5.5	Leadership	2.5												
Writing and legibility	5.0														

Table 15

Number of clusters, major theme, and percentage of references dedicated to each theme in the periods from 2000 to 2022-

2000-2002	%	2003-2005	%	2006-2008	%	2009-2011	%	2012-2014	%	2015-2017	%	2018-2020	%	2021-2022	%
Organizational psychology (Quant methods)	35.2	Organizational psychology (Quant methods)	36.3	Organizational psychology (Quant methods)	36.1	Organizational psychology (Quant methods)	38.6	Strategic management	34.3	Organizational psychology (Quant methods)	36.7	Organizational psychology (Quant methods)	35.1	Strategic management	35.8
Organizational theory	24.0	Strategic management	19.9	Strategic management	27.4	Strategic management	32.1	Organizational psychology (Quant methods)	34.0	Strategic management	30.5	Strategic management	29.9	Organizational psychology (Quant methods)	34.0
Strategic management	16.8	Organizational theory	18.1	Organizational theory	21.8	Organizational theory	24.1	Organizational theory	17.8	Organizational theory	19.2	Organizational theory (Qual methods)	18.0	Organizational theory (Qual methods)	20.1
Organizational structure, embeddedness, and resource dependence	15.8	Critical resources, ownership structure, and strategic alliances	11.6	Upper echelons perspectives and decision making	11.3	Strategic Human Resource Management	5.2	Social networks and learning within the firm	8.9	Upper echelons perspectives and decision making	7.2	Upper echelons perspectives and decision making	10.9	Research Methods	6.3
Upper echelons perspectives	4.6	Upper echelons perspectives	4.9	Strategic Human Resource Management	3.4			Strategic Human Resource Management	5.0	Strategic Human Resource Management	6.4	Strategic Human Resource Management	6.0	Strategic Human Resource Management	3.7
Strategic human resource management	3.5	Strategic Human Resource Management	4.9												
The Functions of the Executive	0.1	Embeddedness and Organization Networks	4.4												

Table 16

Definitions of the eight disciplines used to classify the 225 unique references

Discipline	Definition
Mathematics	The abstract science of number, quantity, and space, either as abstract concepts
Anthropology	The scientific study of human behavioral patterns using field observation methodology
Sociology	The study of the development, structure, and functioning of human groups and organizations
Management[a]	The scientific study of the relationship between internal or external organizational factors and effective administration
Economics	Economics is the scientific study of the behavior and interactions of economic agents
Psychology[b]	The scientific study of the human mind and its functions, especially those affecting behavior in a given context.
Statistics	Statistics deals with the collection, analysis, interpretation, and presentation of masses of numerical data
Biology	The scientific study of living organisms and their evolution

Notes: (a) Qualitative methods such as the Gioia method were classified as management, (b) Quantitative methods and applications that were not purely statistical, such as psychometric scale validations were classified as psychology.

Table 17

Top 20 most important articles in the selected periods, as a function of the frequency with which an article features in the top 5 most cited references of a cluster. Periods within which an article features in the top 5 most cited articles in one of the identified clusters.

Citation	# top 5	Time-periods															
		40 - 60	61 - 70	71 - 80	81 - 85	86 - 90	91 - 93	94 - 96	97 - 99	00 - 02	03 - 05	06 - 08	09 - 11	12 - 14	15 - 17	18 - 20	21 - 22
(Pfeffer & Salancik, 1978)	11				✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	
(March & Simon, 1958)	10		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					
(Barney, 1991)	10							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(DiMaggio & Powell, 1983)	10							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(Jensen & Meckling, 1976)	10					✓		✓	✓	✓	✓	✓	✓		✓	✓	✓
(Cyert & March, 1963)	10		✓		✓		✓	✓		✓		✓	✓	✓	✓	✓	
(Nelson, 1982)	9						✓	✓	✓	✓	✓	✓	✓	✓	✓		
(Thompson, 1967)	8			✓	✓	✓	✓	✓	✓	✓	✓						
(Baron & Kenny, 1986)	8								✓	✓	✓	✓	✓	✓	✓		✓
(Cohen & Levinthal, 1990)	8									✓	✓	✓	✓	✓	✓	✓	✓
(Hambrick & Mason, 1984)	8						✓	✓	✓	✓	✓	✓			✓	✓	
(Hofstede, 1980)	8					✓	✓	✓	✓	✓	✓	✓	✓				
(Huselid, 1995)	8									✓	✓	✓	✓	✓	✓	✓	✓
(Eisenhardt, 1989b)	7										✓	✓	✓	✓	✓	✓	✓
(Aiken, West, & Reno, 1991)	6										✓	✓	✓	✓	✓	✓	
(March, 1991)	6											✓	✓	✓	✓	✓	✓
(Podsakoff, MacKenzie, Lee, & Podsakoff, 2003)	6											✓	✓	✓	✓	✓	✓
(Porter, 1980)	6				✓	✓	✓	✓	✓	✓							
(Williamson, 1975)	6					✓	✓	✓	✓	✓	✓						
(Vroom, 1964)	5		✓	✓	✓	✓	✓										

Note: Management (yellow), economics (light blue), psychology (green), sociology (grey), statistics (dark blue)

Table 18

A summary of the main idea of the 20 most influential articles, as identified in Table 17.

Citation	Summary
Pfeffer and Salancik (1978)	The External Control of Organizations established the resource dependence approach whereby all organizations are dependent on their environment for survival, constraining and controlling the organizations' behavior. To enhance survival organizations can attempt to affect their environment by political means or by forming inter-organizational relationships.
March and Simon (1958)	Organizations established the information processing approach to organizations. They explain organizational phenomena from the perspective of boundedly rational agents who create and maintain simplified subjective representations of objective reality which individuals employ with respect to goals, knowledge, and beliefs about actual and future states of affairs.
Barney (1991)	This article builds on the assumption that strategic resources are stabiliy and heterogeneously distributed across firms, and that these resources can generate sustained competitive advantages. Four empirical indicators of the potential of firm resources to generate sustained competitive advantage are value, rareness, imitability, and substitutability.
(DiMaggio & Powell, 1983)	This article explains what makes organizations so similar. Once a set of organizations emerge as a field, rational actors make their institutions increasingly isomorphic. They describe three isomorphic processes: *coercive* which stems from political influence, *mimetic* which results from imitation under uncertainty, and *normative* resulting from professionalized knowledge.
(Jensen & Meckling, 1976)	This article develops a theory of the ownership structure of the firm based on theories of property rights, agency, and finance. The main thrust of the article is to explain the optimal ownership structure of the firm such as to minimize agency costs, which occur when there is a mismatch between the interests of a principal (the owner) and the agent (the manager).
(Cyert & March, 1963)	This book takes the firm as the basic unit of analysis. It views the firm as a coalition of managers, workers, shareholders, and so on, emphasizes the decision-making process, and attempts to predict behavior with respect to price, output, and resource allocation decisions. Cyert and March propose that firms aim a satisficing rather than maximizing organizational results.
(Nelson, 1982)	Nelson and Winter's work develops an evolutionary theory of the behavior of firms. In this view, firms are motivated by profit, but their actions are not assumed to be profit-maximizing. This theory stresses the tendency of profitable firms to drive others out of business, and firms are seen as both passive, and actively seeking alternatives that affect their environment.
(Thompson, 1967)	Organizations in Action is a multidisciplinary study of the behavior of complex organizations. Central to this perspective is that organizations must handle uncertainty. This perspective considers individual behavior only to the extent that it explains the nature of the organization. Thompson classified organizations according to their technologies and environments.
(Baron & Kenny, 1986)	In this article, Barron and Kenny attempt to distinguish between the properties of moderator and mediator variables. They then delineate conceptual and strategic implications of making use of these distinctions in practical research contexts.
(Cohen & Levinthal, 1990)	In this paper, Cohen and Levinthal argue that a critical factor in a firm's innovative capabilities is to recognize the value of new external information, assimilate it and apply it to commercial ends. They label this capability a firm's absorptive capacity, discuss factors influencing it, and argue that it is in large part a function of the firm's prior level of related knowledge.
(Hambrick & Mason, 1984)	Hambrick and Mason attempt to synthesize a previously fragmented literature into an "upper echelons perspective". Their theory states that organizational outcomes, such as strategic choices and firm performance, are predicted in part by the background characteristics of top management. The organization, then, may be a reflection of its top managers.
(Hofstede, 1980)	In this book, Hofstede proposes four dimensions that can help the reader understand the differences between national cultures. These dimensions are Individualism, Power Distance, Uncertainty Avoidance, and Masculinity.
(Huselid, 1995)	In this study, Huselid evaluates the link between systems of High-Performance Work Practices and firm performance in a national sample of nearly one thousand firms. He finds that these practices have an effect on both intermediate employee outcomes, such as turnover and productivity and also short and long-term measures of financial performance.
(Eisenhardt, 1989b)	In this article, Kathleen Eisenhardt described the process of using case studies to induce theory. She describes the full process from specifying the research question to reaching closure. Her method is highly iterative and tightly linked to data.
(Aiken et al., 1991)	This book provides academics and researchers with tools for estimating, testing, and probing interactions between variables in regression models.
(March, 1991)	In this paper, March contrasts two modes of organizational learning: the exploration of new possibilities and the exploitation of old certainties. It examines complications in allocating resources between the two and argues that adaptive processes, by refining exploitation more rapidly than exploration are likely effective in the short run, but detrimental in the long run.
(Podsakoff et al., 2003)	In this article, the authors examine the extent to which method biases influence behavioral research results. They also identify potential sources of method biases, discuss the cognitive processes through which method biases influence responses to measures, and evaluate different procedural and statistical techniques that are claimed to control method biases.
(Porter, 1980)	In Competitive Strategy, Porter captures the complexity of industry competition in five underlying forces and proposes three generic strategies that aid in the task of strategic positioning: lowest cost, differentiation, and focus. He shows how competitive advantage can be defined in terms of relative cost and relative price, linking it directly to profitability.
(Williamson, 1975)	Williamson analyzes the organization of economic activity within and between markets and hierarchies. He considers the transaction to be the pertinent unit of analysis and defines hierarchical transactions as ones for which a single administrative entity spans both sides of the transaction. He discusses the advantages of the transactional approach by examining three issues: price discrimination, insurance, and vertical integration
(Vroom, 1964)	In this book, Vroom reviews research by psychologists, economists, and sociologists in an attempt to integrate the complex multidisciplinary and multimethod literature on the relationship between motivation and work, bringing concepts such as motive, goal incentive, and attitude to centerstage.

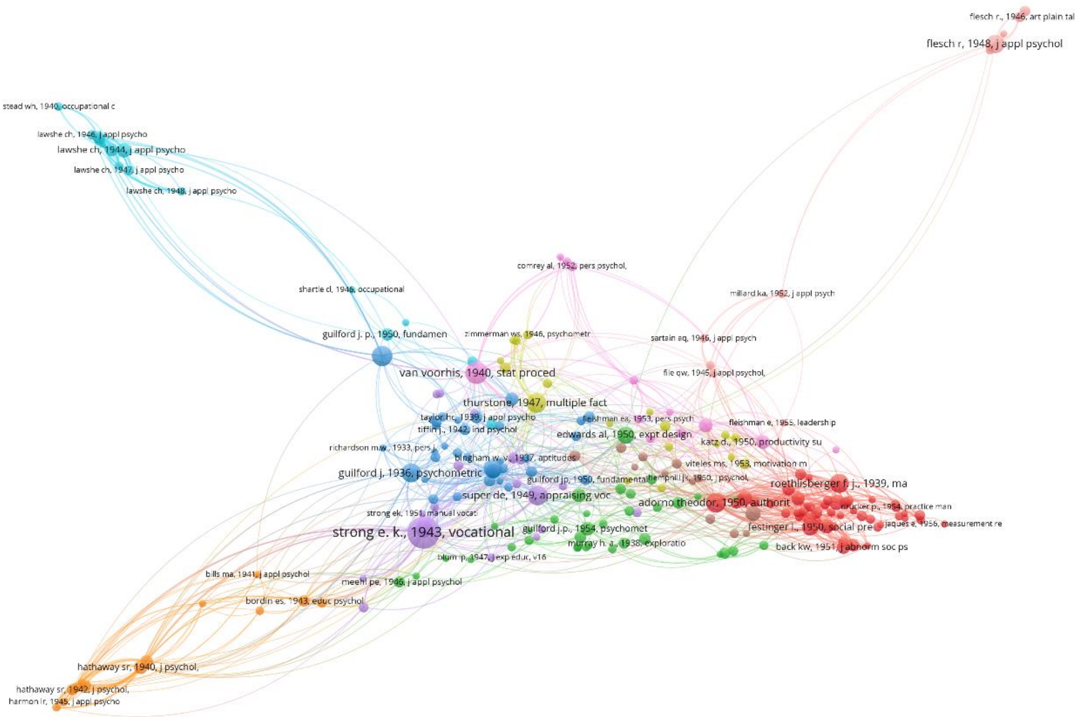
Appendix

Appendix 3

Tables showing the top 5 most cited articles per cluster, and the major theme/theory treated in each cluster for each period.

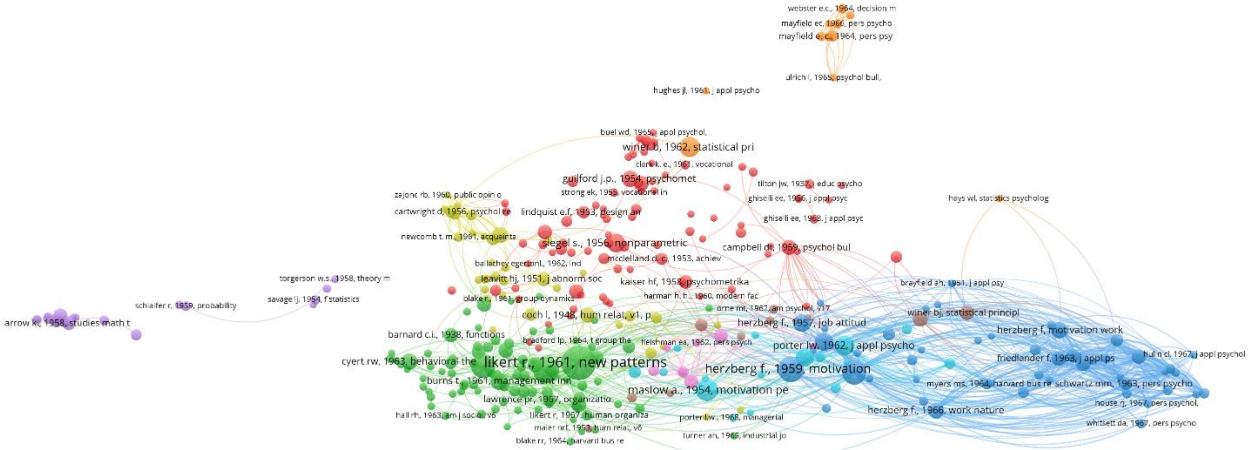
1940-1960

Cluster	Theme / Theory	Major Articles
1	Management	(Adorno, 1950; Coch & French Jr, 1948; Festinger, Schachter, & Back, 1950; Roethlisberger & Dickson, 1939; Whyte. William, 1956)
2	Methods	(Edwards, 1950; Ghiselli, 1954; Guilford, 1954; Kendall, 1948; Lindquist, 1953)
3	Occupational counseling	(Guilford, 1936; McNemar, 1949; Stead et al., 1940; Thorndike, 1949; Viteles & Brief, 1932)
4	Job satisfaction	(Baehr, 1954; Hoppock, 1935; Thurstone, 1947; Tiffin, 1947; Viteles, 1953)
5	Vocational interests	(Bingham, 1937; Darley, 1941; Kuder, 1946; Strong Jr, 1943; Super, 1949)
6	Job characteristics	(Guilford, 1950; Lawshe & Alessi, 1946; Lawshe Jr, 1945; Lawshe Jr & Maleski, 1946; Lawshe Jr & Satter, 1944)
7	Multiphasic personality inventory	(Bordin, 1943; Hathaway & McKinley, 1942; Hathaway & McKinley, 1940; Hathaway & McKinley, 1943; McKinley & Hathaway, 1940)
8	Group psychology	(Bales, 1950; Guilford, 1942; Krech, 1948; Newcomb & Charters Jr, 1950; Thurstone, 1929)
9	Leadership	(Comrey, Pfiffner, & Beem, 1952; Fleishman, 1951; Katz, Maccoby, Gurin, & Floor, 1951; Katz, Maccoby, & Morse, 1950; Peters & Van Voorhis, 1940)
10	Writing and legibility	(File, 1945; File & Remmers, 1946; Flesch, 1946; Flesch, 1948; Flesch, 1949)



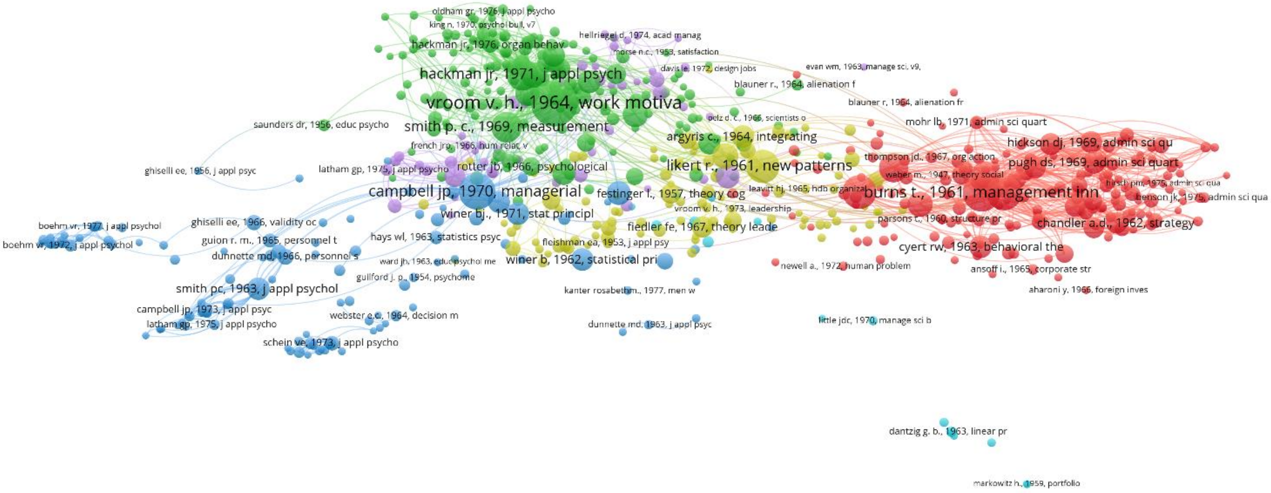
1961-1970

Cluster	Theme / Theory	Major Articles
1	Methods	(Campbell & Fiske, 1959; Guilford, 1954; Osgood, Suci, & Tannenbaum, 1957; Siegel, 1956; Strong Jr, 1943)
2	“New management”, sociology within the organization	(Burns & Stalker, 1961; Cyert & March, 1963; Likert, 1961; March & Simon, 1958; McGregor & Cutcher-Gershenfeld, 1960)
3	Work motivation	(Brayfield & Crockett, 1955; Herzberg, Mausner, & Snyderman, 1959; Herzberg, Mausnes, Peterson, & Capwell, 1957; Herzberg, 1966; Vroom, 1964)
4	Group and interpersonal psychology	(Cartwright & Harary, 1956; Coch & French Jr, 1948; Festinger, 1957; Heider, 1958; Leavitt, 1951)
5	Dynamic and linear programming (optimization under constraints)	(Arrow, Karlin, & Scarf, 1958; Bellman, 1957; Charnes & Cooper, 1961; Holt, 1960; Markowitz, 1959)
6	Maslow need satisfaction from management	(Maslow & Frager, 1954; Maslow, 1943; Porter, 1961; Porter, 1962, 1963)
7	Job interviews	(Mayfield, 1964; Mayfield & Carlson, 1966; Wagner, 1949; Webster, 1964; Winer, 1962)
8	Factors influencing productivity	(Georgopoulos, Mahoney, & Jones Jr, 1957; Maier, 1958; Orne, 1962; Vroom, 1964; Winer, 1962)
9	Leadership	(Fleishman, 1953a; Fleishman, 1953b; Fleishman & Harris, 1962; Fleishman, Harris, & Burtt, 1955; Fleishman & Peters, 1962)



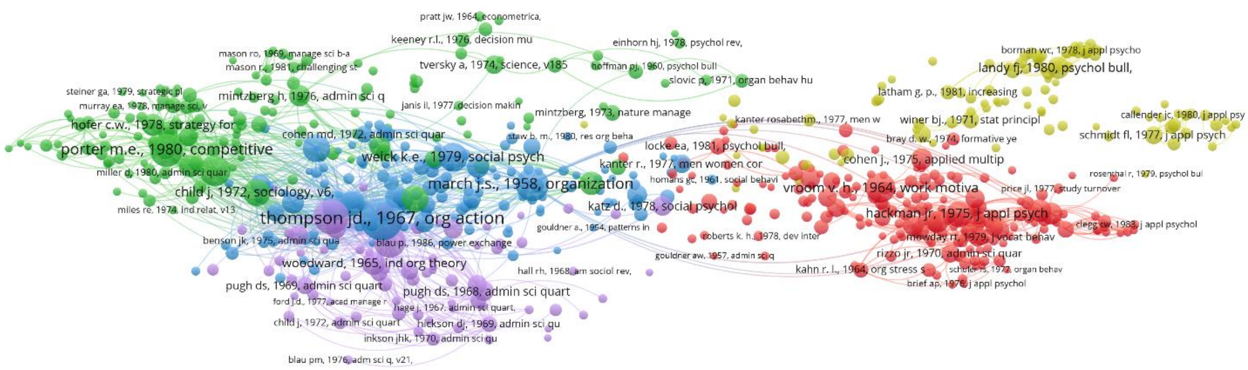
1971-1980

Cluster	Theme / Theory	Major Articles
1	Organizational theory	(Burns & Stalker, 1961; Katz & Kahn, 1966; Lawrence & Lorsch, 1967; March & Simon, 1958; Thompson, 1967)
2	Motivation and Job satisfaction	(Hackman & Lawler, 1971; Hackman & Oldham, 1975; Maslow & Frager, 1954; Smith, Kendall, & Hulin, 1969; Vroom, 1964)
3	Methods (in management)	(Campbell, Dunnette, Lawler, & Weick Jr, 1970; Nunnally & Bernstein, 1967; Smith & Kendall, 1963; Winer, 1971; Winer, 1962)
4	"New management"	(Argyris, 1964; Festinger, 1957; Fiedler, 1967; Likert, 1961; McGregor & Cutcher-Gershenfeld, 1960)
5	Human performance (in the organization)	(Heider, 1958; Likert, 1967; Litwin & Stringer, 1968; Locke, 1968; Porter & Lawler, 1965)
6	Managerial decision making	(Adorno, 1950; Dantzig, 1963; Little, 1970; Rokeach, 1960; Schroder, Driver, & Streufert, 1967)



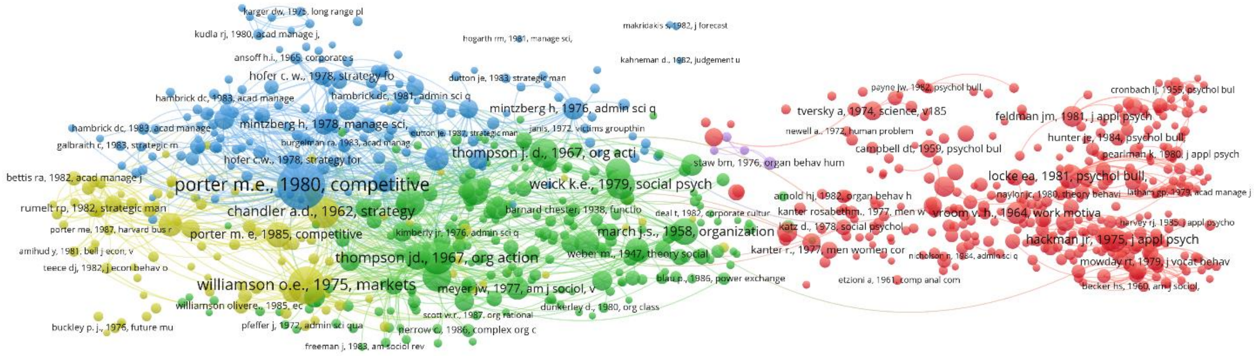
1981-1985

Cluster	Theme / Theory	Major Articles
1	Motivation and job satisfaction	(Hackman & Lawler, 1971; Hackman & Oldham, 1975; Mobley, Griffeth, Hand, & Meglino, 1979; Porter & Steers, 1973; Vroom, 1964)
2	Strategic management	(Chandler Jr, 1962; Child, 1972; March & Simon, 1958; Porter, 1980; Thompson, 1967)
3	Organizational theory	(Cyert & March, 1963; March & Simon, 1958; Pfeffer & Salancik, 1978; Thompson, 1967; Weick, 1979)
4	Methods (in management)	(Campbell et al., 1970; Feldman, 1981; Landy & Farr, 1980; Schmidt & Hunter, 1977; Winer, 1971)
5	Organizational design	(Burns & Stalker, 1961; Galbraith, 1973; Katz & Kahn, 1978; Lawrence & Lorsch, 1967; Woodward, 1965)



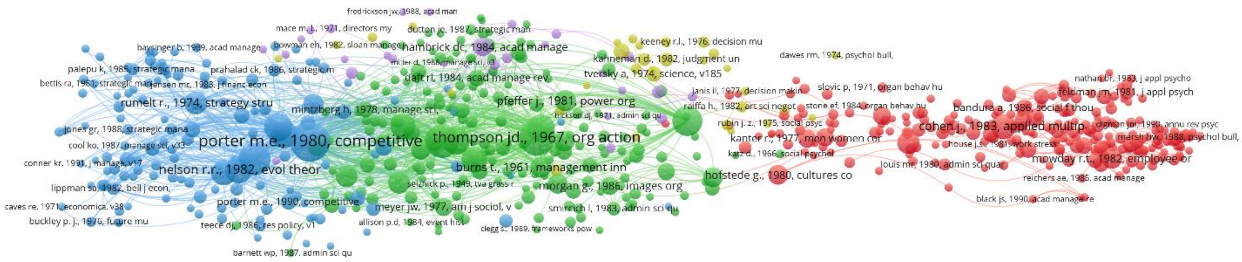
1986-1990

Cluster	Theme / Theory	Major Articles
1	Motivation and Job satisfaction	(Feldman, 1981; Hackman & Oldham, 1975; Hofstede, 1980; Locke, Shaw, Saari, & Latham, 1981; Vroom, 1964)
2	Organizational theory	(Burns & Stalker, 1961; March & Simon, 1958; Pfeffer & Salancik, 1978; Thompson, 1967; Weick, 1979)
3	Strategic management	(Lawrence & Lorsch, 1967; Miles, Snow, Meyer, & Coleman Jr, 1978; Mintzberg, 1978; Mintzberg, Raisinghani, & Theoret, 1976; Porter, 1980)
4	Organizational structure	(Chandler Jr, 1962; Jensen & Meckling, 1976; Porter, 1985; Rumelt, 1974; Williamson, 1975)
5	The "trapped administrator"	(Fox & Staw, 1979; Staw, 1976; Staw & Fox, 1977; Staw & Ross, 1978)



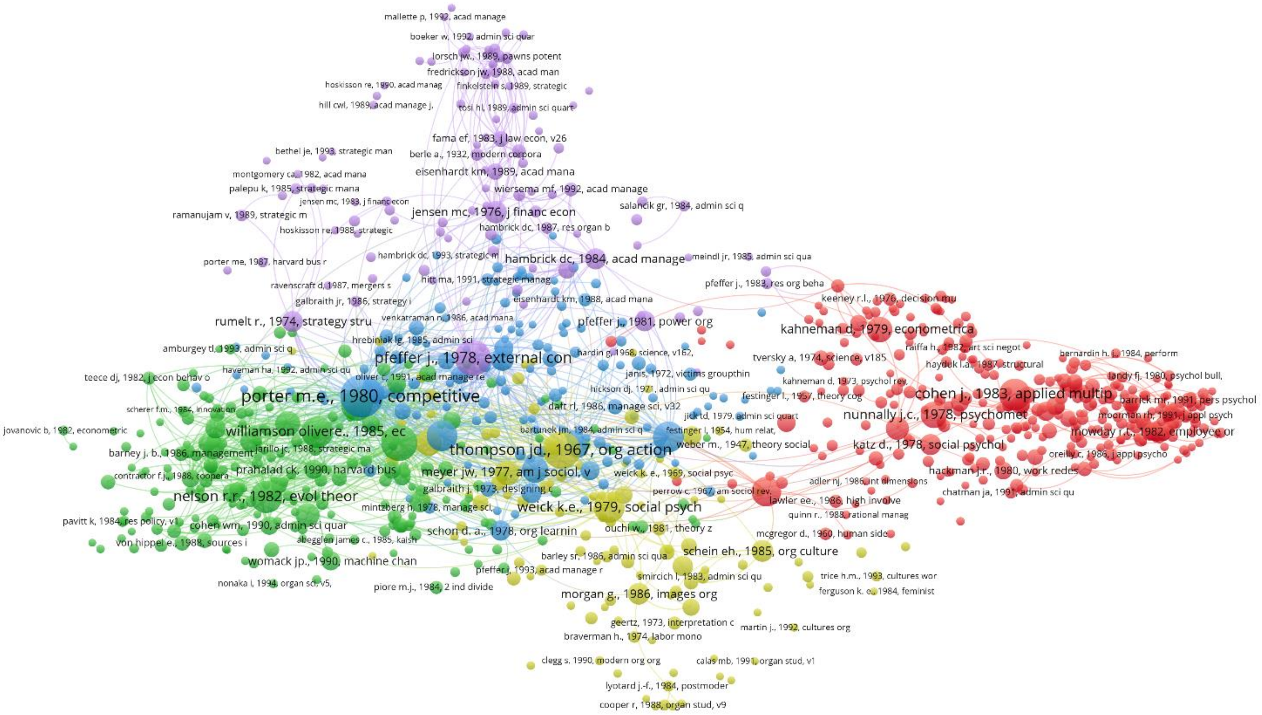
1991-1993

Cluster	Theme / Theory	Major Articles
1	Organizational psychology	(Cohen, Cohen, West, & Aiken, 1983; Hofstede, 1980; Mowday, Porter, & Steers, 1982; Nunnally, 1978; Vroom, 1964)
2	Organizational theory	(Cyert & March, 1963; March & Simon, 1958; Pfeffer & Salancik, 1978; Thompson, 1967; Weick, 1979)
3	Strategic management & Organizational structure	(Chandler Jr, 1962; Nelson, 1982; Porter, 1980, 1985; Williamson, 1975)
4	Heuristics, Biases, and Decision making	(Axelrod, 1984; Kahneman, Slovic, Slovic, & Tversky, 1982; Kahneman & Tversky, 1979; Tversky & Kahneman, 1974, 1981)
5	Upper echelons perspectives	(Fama & Jensen, 1983; Gupta & Govindarajan, 1984; Hambrick & Mason, 1984; Mintzberg, 1973; Mintzberg, 1983)



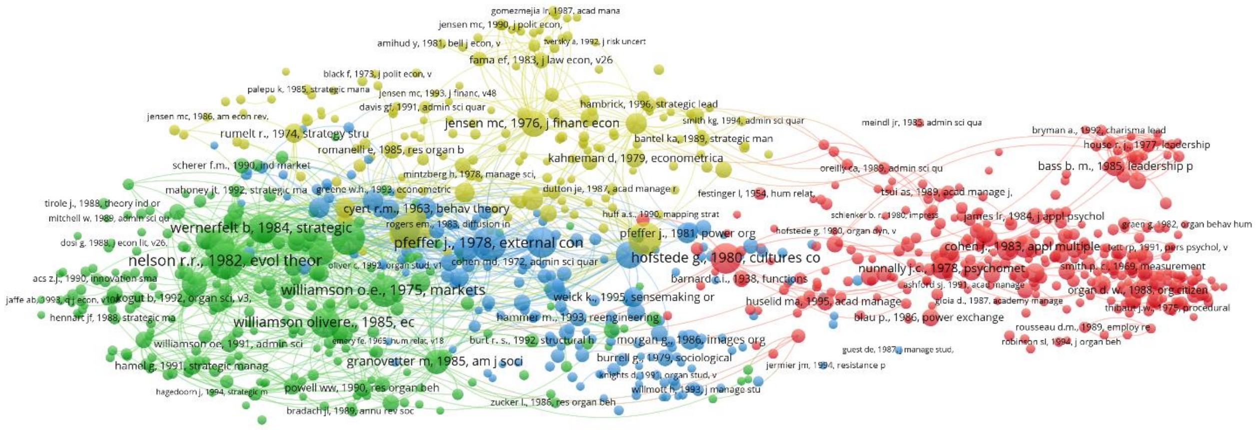
1994-1996

Cluster	Theme / Theory	Major Articles
1	Organizational psychology	(Cohen et al., 1983; Hofstede, 1980; Kahneman & Tversky, 1979; Kanter, 1977; Nunnally, 1978)
2	Organizational structure	(Barney, 1991; Nelson, 1982; Porter, 1985; Williamson, 1975, 1985)
3	Strategic Management	(Chandler Jr, 1962; Cyert & March, 1963; Lawrence & Lorsch, 1967; March & Simon, 1958; Porter, 1980)
4	Organizational Theory	(DiMaggio & Powell, 1983; Meyer & Rowan, 1977; Morgan, 1986; Thompson, 1967; Weick, 1979)
5	Upper echelons perspectives	(Hambrick & Mason, 1984; Jensen & Meckling, 1976; Pfeffer & Pfeffer, 1981; Pfeffer & Salancik, 1978; Rumelt, 1974)



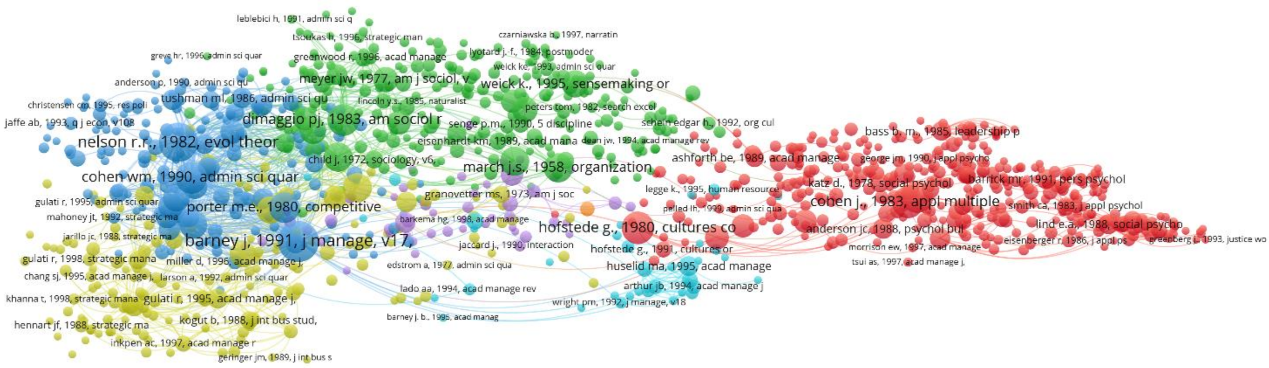
1997-1999

Cluster	Theme / Theory	Major Articles
1	Organizational psychology	(Baron & Kenny, 1986; Bass & Bass, 1985; Cohen et al., 1983; Hofstede, 1980; Nunnally, 1978)
2	Organizational structure and resource dependence	(Barney, 1991; Nelson, 1982; Wernerfelt, 1984; Williamson, 1975, 1985)
3	Organizational theory	(DiMaggio & Powell, 1983; Meyer & Rowan, 1977; Pfeffer & Salancik, 1978; Thompson, 1967; Weick, 1979)
4	Strategic management / Upper echelons perspectives	(Hambrick & Mason, 1984; Jensen & Meckling, 1976; March & Simon, 1958; Porter, 1980, 1985)



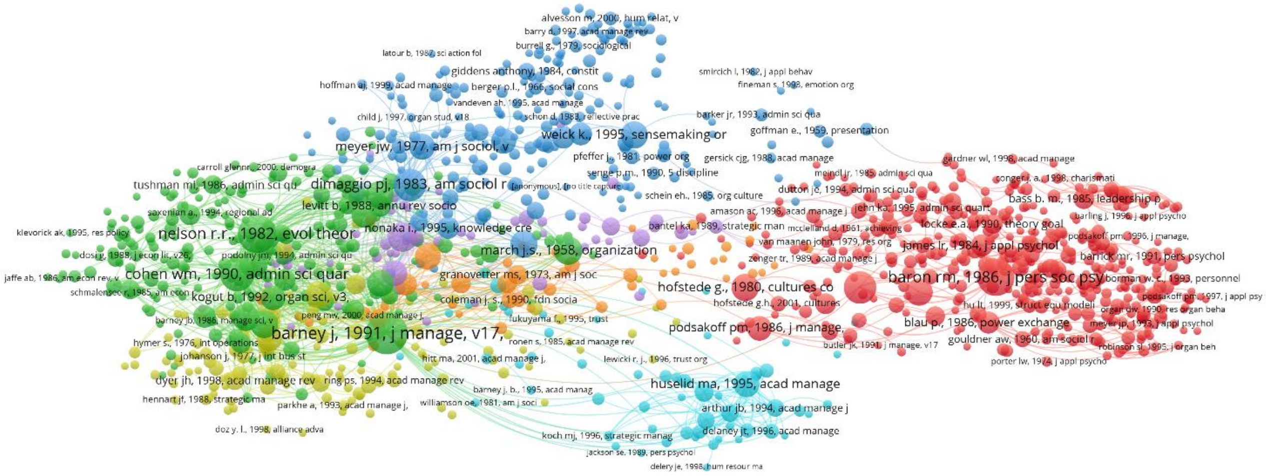
2000-2002

Cluster	Theme / Theory	Major Articles
1	Organizational psychology (Quant methods)	(Baron & Kenny, 1986; Cohen et al., 1983; Hofstede, 1980; Nunnally, 1978; Podsakoff & Organ, 1986)
2	Organizational theory	(Cyert & March, 1963; DiMaggio & Powell, 1983; March & Simon, 1958; Thompson, 1967; Weick, 1995)
3	Strategic Management	(Barney, 1991; Cohen & Levinthal, 1990; Nelson, 1982; Porter, 1980; Wernerfelt, 1984)
4	Organizational structure, embeddedness, and resource dependence	(Burt, 1992; Granovetter, 1985; Pfeffer & Salancik, 1978; Williamson, 1975, 1985)
5	Upper echelons perspectives	(Bantel & Jackson, 1989; Hambrick & Finkelstein, 1996; Hambrick & Mason, 1984; Jensen & Meckling, 1976; Wiersema & Bantel, 1992)
6	Strategic human resource management	(Arthur, 1994; Becker & Gerhart, 1996; Huselid, 1995; MacDuffie, 1995; Miles et al., 1978)
7	The Functions of the Executive	(Barnard, 1938)



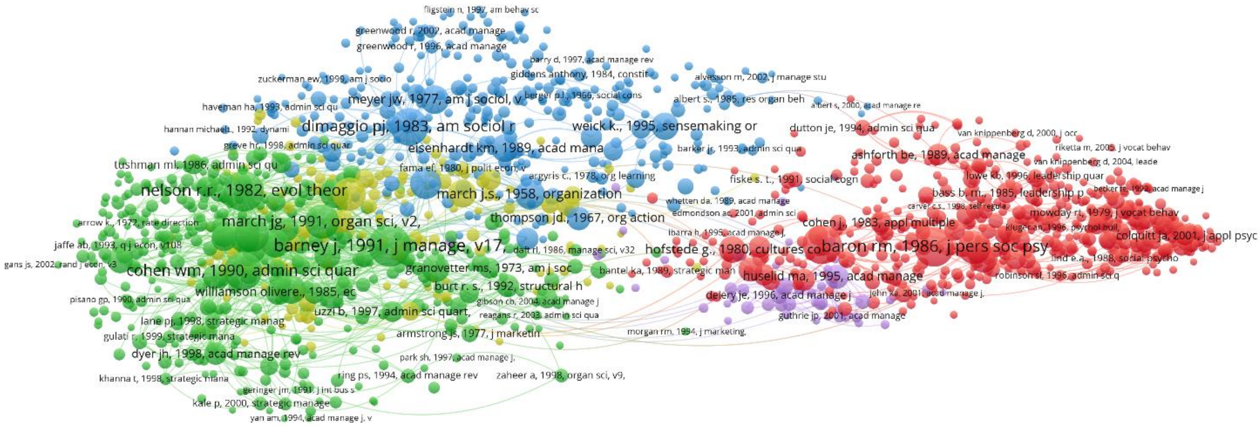
2003-2005

Cluster	Theme / Theory	Major Articles
1	Organizational psychology (Quant methods)	(Aiken et al., 1991; Baron & Kenny, 1986; Cohen et al., 1983; Hofstede, 1980; Podsakoff & Organ, 1986)
2	Strategic management	(Barney, 1991; Cohen & Levinthal, 1990; March & Simon, 1958; Nelson, 1982; Teece, Pisano, & Shuen, 1997)
3	Organizational theory	(DiMaggio & Powell, 1983; Eisenhardt, 1989a; Meyer & Rowan, 1977; Thompson, 1967; Weick, 1995)
4	Critical resources, ownership structure, and strategic alliances	(Dyer & Singh, 1998; Gulati, 1995; Szulanski, 1996; Williamson, 1975, 1985)
5	Upper echelons perspectives	(Eisenhardt, 1989c; Hambrick & Finkelstein, 1996; Hambrick & Mason, 1984; Jensen & Meckling, 1976; Pfeffer & Salancik, 1978)
6	Strategic Human Resource Management	(Arthur, 1994; Becker & Gerhart, 1996; Delery & Doty, 1996; Huselid, 1995; MacDuffie, 1995)
7	Embeddedness and Organization Networks	(Burt, 1992; Granovetter, 1985; Granovetter, 1973; Nahapiet & Ghoshal, 1998; Uzzi, 1997)



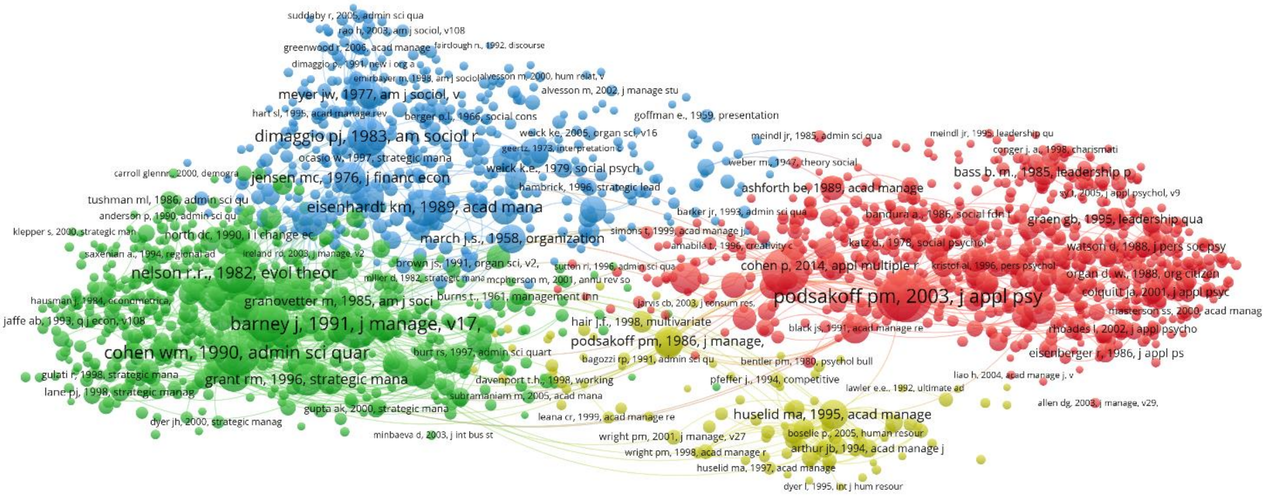
2006-2008

Cluster	Theme / Theory	Major Articles
1	Organizational psychology (Quant methods)	(Aiken et al., 1991; Baron & Kenny, 1986; Blau, 1986; Hofstede, 1980; Podsakoff et al., 2003)
2	Strategic management	(Barney, 1991; Cohen & Levinthal, 1990; March, 1991; Nelson, 1982; Teece et al., 1997)
3	Organizational theory	(DiMaggio & Powell, 1983; Eisenhardt, 1989b; March & Simon, 1958; Meyer & Rowan, 1977; Pfeffer & Salancik, 1978)
4	Upper echelons perspectives and decision making	(Cyert & March, 1963; Hambrick & Mason, 1984; Jensen & Meckling, 1976; Kahneman & Tversky, 1979; Levitt & March, 1988)
5	Strategic Human Resource Management	(Arthur, 1994; Becker & Gerhart, 1996; Delery & Doty, 1996; Huselid, 1995; MacDuffie, 1995)



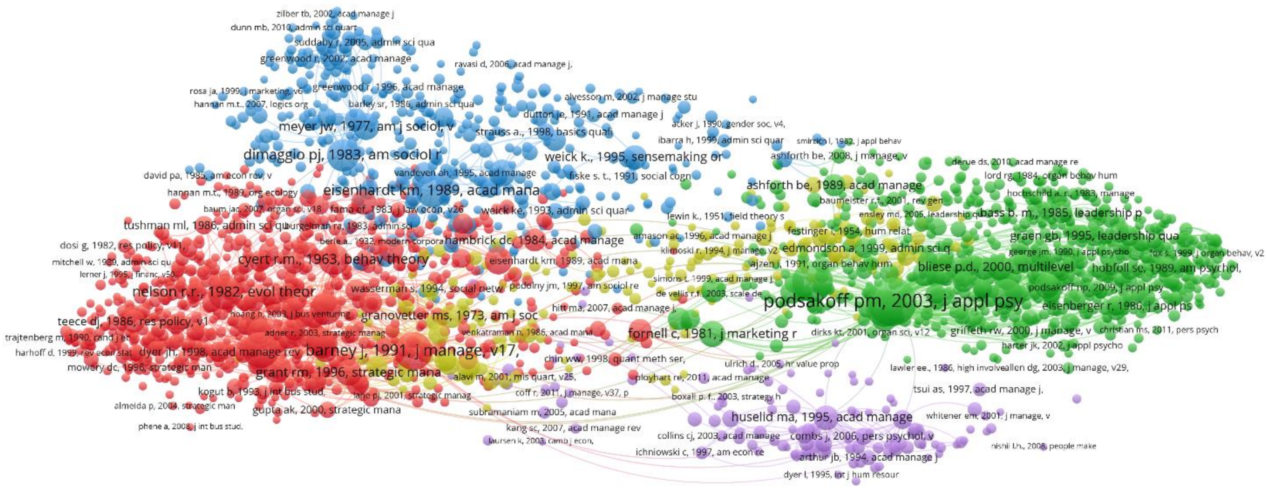
2009-2011

Cluster	Theme / Theory	Major Articles
1	Organizational psychology (Quant methods)	(Aiken et al., 1991; Baron & Kenny, 1986; Blau, 1986; Hofstede, 1980; Podsakoff et al., 2003)
2	Strategic management	(Barney, 1991; Cohen & Levinthal, 1990; March, 1991; Nelson, 1982; Teece et al., 1997)
3	Organizational theory	(Cyert & March, 1963; DiMaggio & Powell, 1983; Eisenhardt, 1989b; Jensen & Meckling, 1976; Pfeffer & Salancik, 1978)
4	Strategic Human Resource Management	(Fornell & Larcker, 1981; Hair, Black, Babin, Anderson, & Tatham, 1998; Huselid, 1995; MacDuffie, 1995; Podsakoff & Organ, 1986)



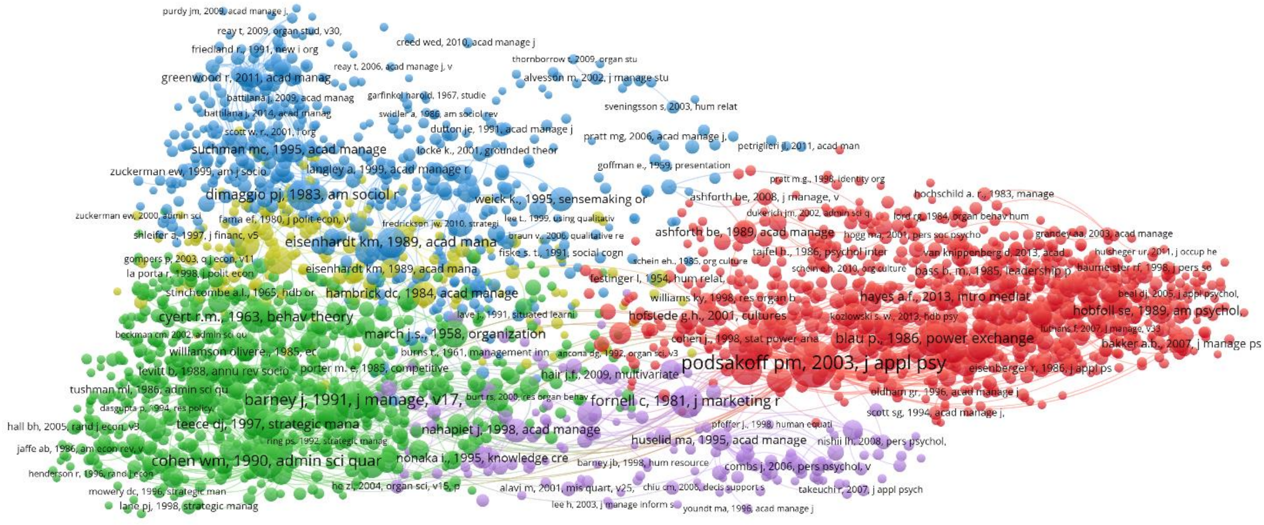
2012-2014

Cluster	Theme / Theory	Major Articles
1	Strategic management	(Barney, 1991; Cohen & Levinthal, 1990; Cyert & March, 1963; March, 1991; Nelson, 1982)
2	Organizational psychology (Quant methods)	(Aiken et al., 1991; Baron & Kenny, 1986; Blau, 1986; Fornell & Larcker, 1981; Podsakoff et al., 2003)
3	Organizational theory	(DiMaggio & Powell, 1983; Eisenhardt, 1989b; Meyer & Rowan, 1977; Miles, Huberman, Huberman, & Huberman, 1994; Weick, 1995)
4	Social networks and learning within the firm	(Coleman, 1988; Granovetter, 1973; Hansen, 1999; Nahapiet & Ghoshal, 1998; Nonaka, o Nonaka, Ikujiro, & Takeuchi, 1995)
5	Strategic Human Resource Management	(Bowen & Ostroff, 2004; Combs, Liu, Hall, & Ketchen, 2006; Delery & Doty, 1996; Huselid, 1995; Lepak & Snell, 1999)



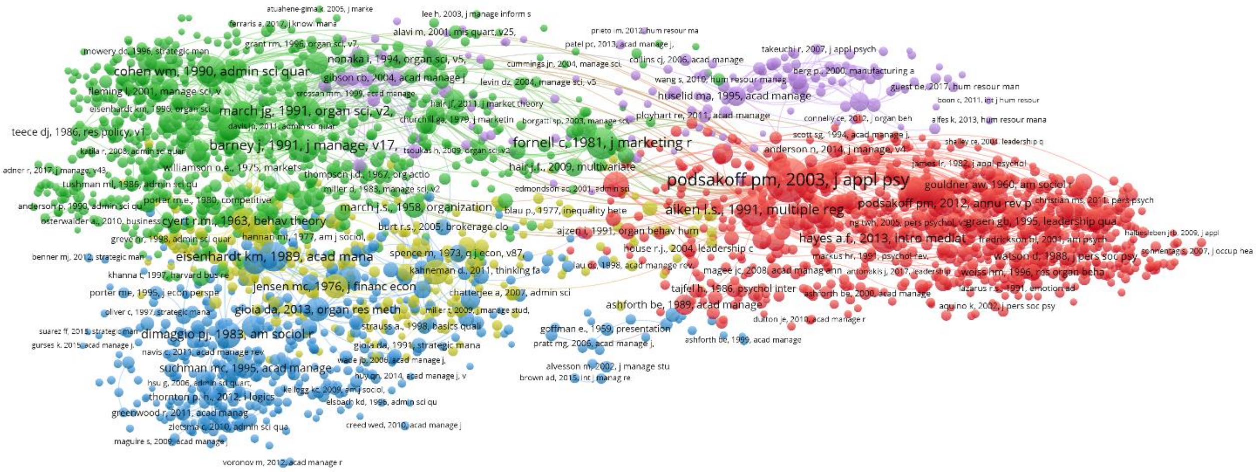
2015-2017

Cluster	Theme / Theory	Major Articles
1	Organizational psychology (Quant methods)	(Aiken et al., 1991; Baron & Kenny, 1986; Blau, 1986; Hu & Bentler, 1999; Podsakoff et al., 2003)
2	Strategic management	(Barney, 1991; Cohen & Levinthal, 1990; Cyert & March, 1963; March, 1991; Nelson, 1982)
3	Organizational theory	(DiMaggio & Powell, 1983; Eisenhardt, 1989b; Eisenhardt & Graebner, 2007; Miles et al., 1994; Suchman, 1995)
4	Upper echelons perspectives and decision making	(Eisenhardt, 1989a; Finkelstein, Hambrick, & Cannella, 2009; Hambrick & Mason, 1984; Jensen & Meckling, 1976; Pfeffer & Salancik, 1978)
5	Strategic Human Resource Management	(Fornell & Larcker, 1981; Huselid, 1995; Nahapiet & Ghoshal, 1998; Nonaka et al., 1995; Podsakoff & Organ, 1986)



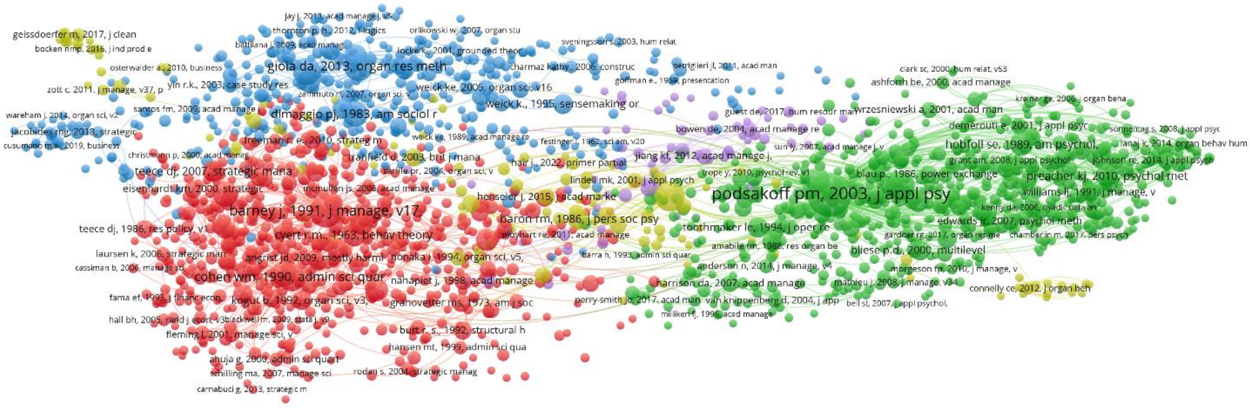
2018-2020

Cluster	Theme / Theory	Major Articles
1	Organizational psychology (Quant methods)	(Aiken et al., 1991; Blau, 1986; Hayes, 2013; Podsakoff et al., 2003; Podsakoff, MacKenzie, & Podsakoff, 2012)
2	Strategic management	(Barney, 1991; Cohen & Levinthal, 1990; Fornell & Larcker, 1981; March, 1991; Teece et al., 1997)
3	Organizational theory (Qual methods)	(DiMaggio & Powell, 1983; Eisenhardt, 1989b; Eisenhardt & Graebner, 2007; Gioia, Corley, & Hamilton, 2013; Miles et al., 1994)
4	Upper echelons perspectives and decision making	(Cyert & March, 1963; Hambrick & Mason, 1984; Jensen & Meckling, 1976; Ocasio, 1997; Pfeffer & Salancik, 1978)
5	Strategic Human Resource Management	(Bowen & Ostroff, 2004; Combs et al., 2006; Gibson & Birkinshaw, 2004; Huselid, 1995; Jiang, Lepak, Hu, & Baer, 2012)



2021-2022

Cluster	Theme / Theory	Major Articles
1	Strategic management	(Barney, 1991; Cohen & Levinthal, 1990; Jensen & Meckling, 1976; March, 1991; Teece et al., 1997)
2	Organizational psychology (Quant methods)	(Bolin, 2014; Hobfoll, 1989; Hu & Bentler, 1999; Podsakoff et al., 2003; Podsakoff et al., 2012)
3	Organizational theory (Qual methods)	(DiMaggio & Powell, 1983; Eisenhardt, 1989b; Eisenhardt & Graebner, 2007; Gioia et al., 2013; Langley, 1999)
4	Research methods	(Armstrong & Overton, 1977; Baron & Kenny, 1986; Fornell & Larcker, 1981; Podsakoff & Organ, 1986; Tranfield, Denyer, & Smart, 2003)
5	Strategic Human Resource Management	(Bowen & Ostroff, 2004; Combs et al., 2006; Huselid, 1995; Jiang et al., 2012; Lepak & Snell, 1999)



Appendix References

Adorno, T. W. (1950). The authoritarian personality. *Social studies series*.

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*: sage.

Argyris, C. (1964). Integrating the individual and the organization.

Armstrong, J. S., & Overton, T. S. (1977). Estimating nonresponse bias in mail surveys. *Journal of Marketing Research, 14*(3), 396-402.

Arrow, K. J., Karlin, S., & Scarf, H. E. (1958). Studies in the mathematical theory of inventory and production.

Arthur, J. B. (1994). Effects of human resource systems on manufacturing performance and turnover. *Academy of management journal, 37*(3), 670-687.

Axelrod, R. (1984). *Evolution Of Cooperation*: Basic Books.

Baehr, M. E. (1954). A factorial study of the SRA Employee Inventory. *Personnel Psychology*.

Bales, R. F. (1950). Interaction process analysis; a method for the study of small groups.

Bantel, K. A., & Jackson, S. E. (1989). Top management and innovations in banking: Does the composition of the top team make a difference? *Strategic Management Journal, 10*(S1), 107-124.

Barnard, C. (1938). The functions of the executive.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99-120.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173.

Bass, B. M., & Bass, B. M. B. (1985). *Leadership and Performance Beyond Expectations*: Free Press.

Becker, B., & Gerhart, B. (1996). The impact of human resource management on organizational performance: Progress and prospects. *Academy of management journal, 39*(4), 779-801.

Bellman, R. (1957). Dynamic programming. *Princeton, USA: Princeton University Press, 1*(2), 3.

Bingham, W. V. (1937). Aptitudes and aptitude testing.

Blau, P. M. (1986). *Exchange and power in social life*: Routledge.

Bolin, J. H. (2014). Introduction to mediation, moderation, and conditional process analysis: a regression-based approach. In: JSTOR.

Bordin, E. S. (1943). A Theory of Vocational Interests as Dynamic Phenomena'. *Educational and Psychological Measurement, 3*(1), 46-62.

Bowen, D. E., & Ostroff, C. (2004). Understanding HRM–firm performance linkages: The role of the "strength" of the HRM system. *Academy of Management review, 29*(2), 203-221.

Brayfield, A. H., & Crockett, W. H. (1955). Employee attitudes and employee performance. *Psychological Bulletin, 52*(5), 396.

Burns, T., & Stalker, G. M. (1961). The management of innovation. London. *Tavistock Publishing. Cited in Hurley, RF and Hult, GTM (1998). Innovation, Market Orientation, and Organisational Learning: An Integration and Empirical Examination. Journal of Marketing, 62*, 42-54.

Burt, R. S. (1992). Structural holes. In *structural holes*: Harvard university press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81.

Campbell, J. P., Dunnette, M. D., Lawler, E. E., & Weick Jr, K. E. (1970). *Managerial Behavior, Performance and Effectiveness*: McGraw-Hill Inc.,US.

Cartwright, D., & Harary, F. (1956). Structural balance: a generalization of Heider's theory. *Psychological review, 63*(5), 277-293. doi:10.1037/h0046049

Chandler Jr, A. D. (1962). *Strategy and structure: chapters in the history of the industrial enterprise*. Oxford, England: M.I.T. Press.

Charnes, A., & Cooper, W. W. (1961). *Management Models and the Industrial Applications of Linear Programming*: John Wiley, New York.

Child, J. (1972). Organizational structure, environment and performance: The role of strategic choice. *sociology, 6*(1), 1-22.
Coch, L., & French Jr, J. R. (1948). Overcoming resistance to change. *Human relations, 1*(4), 512-532.
Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (1983). Applied multiple regression. *Correlation Analysis for the Behavioral Sciences, 2*.
Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 128-152.
Coleman, J. S. (1988). Social capital in the creation of human capital. *American journal of sociology, 94*, S95-S120.
Combs, J., Liu, Y., Hall, A., & Ketchen, D. (2006). How much do high-performance work practices matter? A meta-analysis of their effects on organizational performance. *Personnel Psychology, 59*(3), 501-528.
Comrey, A. L., Pfiffner, J. M., & Beem, H. P. (1952). Factors influencing organizational effectiveness. I. The US Forest Survey. *Personnel Psychology*.
Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm* (Vol. 2): Englewood Cliffs, NJ.
Dantzig, G. (1963). *Linear Programming and Extensions*: Princeton UniversityPress, Princeton, NJ, .
Darley, J. G. (1941). Clinical aspects and interpretation of the Strong Vocational Interest Blank.
Delery, J. E., & Doty, D. H. (1996). Modes of theorizing in strategic human resource management: Tests of universalistic, contingency, and configurational performance predictions. *Academy of management journal, 39*(4), 802-835.
DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American sociological review*, 147-160.
Dyer, J. H., & Singh, H. (1998). The relational view: Cooperative strategy and sources of interorganizational competitive advantage. *Academy of Management review, 23*(4), 660-679.
Edwards, A. L. (1950). Experimental design in psychological research. New York: Rinehart and Company. *Inc., i960*.
Eisenhardt, K. M. (1989a). Agency theory: An assessment and review. *Academy of Management review, 14*(1), 57-74.
Eisenhardt, K. M. (1989b). Building theories from case study research. *Academy of Management review, 14*(4), 532-550.
Eisenhardt, K. M. (1989c). Making fast strategic decisions in high-velocity environments. *Academy of management journal, 32*(3), 543-576.
Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of management journal, 50*(1), 25-32.
Fama, E. F., & Jensen, M. C. (1983). Agency problems and residual claims. *The journal of law and Economics, 26*(2), 327-349.
Feldman, D. C. (1981). The multiple socialization of organization members. *Academy of Management review, 6*(2), 309-318.
Festinger, L. (1957). A theory of cognitive dissonance.
Festinger, L., Schachter, S., & Back, K. (1950). Social pressures in informal groups; a study of human factors in housing.
Fiedler, F. E. (1967). A theory of Leadership Effectiveness. McGraw-Hill Series in Management.
File, Q. W. (1945). The measurement of supervisory quality in industry. *Journal of Applied Psychology, 29*(5), 323.
File, Q. W., & Remmers, H. H. (1946). Studies in supervisory evaluation. *Journal of Applied Psychology, 30*(5), 421.
Finkelstein, S., Hambrick, D. C., & Cannella, A. A. (2009). *Strategic Leadership: Theory and Research on Executives, Top Management Teams, and Boards*: Oxford University Press.
Fleishman, E. A. (1951). *" Leadership Climate" and Supervisory Behavior: A Study of the Leadership Role of the Foreman in an Industrial Situation*: Ohio State University Personnel Research Board.
Fleishman, E. A. (1953a). The description of supervisory behavior. *Journal of Applied Psychology, 37*(1), 1.
Fleishman, E. A. (1953b). Leadership Climate, Human Relations Training, and Supervisory Behavior. *Personnel Psychology, 6*(2), 205-222. doi:https://doi.org/10.1111/j.1744-6570.1953.tb01040.x

Fleishman, E. A., & Harris, E. F. (1962). Patterns of leadership behavior related to employee grievances and turnover. *Personnel Psychology*.

Fleishman, E. A., Harris, E. F., & Burtt, H. E. (1955). Leadership and supervision in industry; an evaluation of a supervisory training program. *Ohio State University. Bureau of Educational Research Monograph, 33*, xiii, 110-xiii, 110.

Fleishman, E. A., & Peters, D. R. (1962). Interpersonal values, leadership attitudes and managerial" success.". *Personnel Psychology*.

Flesch, R. (1946). The art of plain talk.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*(3), 221.

Flesch, R. F. (1949). Art of readable writing.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50.

Fox, F. V., & Staw, B. M. (1979). The trapped administrator: Effects of job insecurity and policy resistance upon commitment to a course of action. *Administrative Science Quarterly*, 449-471.

Galbraith, J. (1973). Designing complex organizations. *Reading, Mass*.

Georgopoulos, B. S., Mahoney, G. M., & Jones Jr, N. W. (1957). A path-goal approach to productivity. *Journal of Applied Psychology, 41*(6), 345-353. doi:10.1037/h0048473

Ghiselli, E. E. (1954). The forced-choice technique in self-description. *Personnel Psychology*.

Gibson, C. B., & Birkinshaw, J. (2004). The antecedents, consequences, and mediating role of organizational ambidexterity. *Academy of management journal, 47*(2), 209-226.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods, 16*(1), 15-31.

Granovetter, M. (1985). Economic Action and Social Structure: The Problem of~ mbeddednessl. *American journal of sociology, 91*(3), 481-510.

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology, 78*(6), 1360-1380.

Guilford, J. (1936). Psychometric Methods. In: New York: McGraw-Hill.

Guilford, J. P. (1942). Fundamental Statistics in Psychology and Education: By JP Guilford. In: McGraw-Hill Book Company.

Guilford, J. P. (1950). Fundamental statistics in psychology and education.

Guilford, J. P. (1954). Psychometric methods.

Gulati, R. (1995). Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances. *Academy of management journal, 38*(1), 85-112.

Gupta, A. K., & Govindarajan, V. (1984). Business unit strategy, managerial characteristics, and business unit effectiveness at strategy implementation. *Academy of management journal, 27*(1), 25-41.

Hackman, J. R., & Lawler, E. E. (1971). Employee reactions to job characteristics. *Journal of Applied Psychology, 55*(3), 259.

Hackman, J. R., & Oldham, G. R. (1975). Development of Job Diagnostic Survey. *Journal of Applied Psychology, 60*(2), 159-170. doi:10.1037/h0076546

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate data analysis* (Vol. 5): Prentice hall Upper Saddle River, NJ.

Hambrick, D. C., & Finkelstein, S. (1996). *Strategic Leadership: Top Executives and Their Effects on Organizations*: West Publishing Company.

Hambrick, D. C., & Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management review, 9*(2), 193-206.

Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly, 44*(1), 82-111.

Hathaway, S. R., & McKinley, J. (1942). A multiphasic personality schedule (Minnesota): III. The measurement of symptomatic depression. *The Journal of Psychology, 14*(1), 73-84.

Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *The Journal of Psychology, 10*(2), 249-254.

Hathaway, S. R., & McKinley, J. C. (1943). The Minnesota multiphasic personality inventory, Rev. ed., 2nd printing.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY, US: Guilford Press.

Heider, F. (1958). The psychology of interpersonal relations.
Herzberg, F., Mausner, B., & Snyderman, B. (1959). The motivation to work.
Herzberg, F., Mausnes, B., Peterson, R. O., & Capwell, D. F. (1957). Job attitudes; review of research and opinion.
Herzberg, F. I. (1966). Work and the Nature of Man.
Hobfoll, S. E. (1989). Conservation of resources: a new attempt at conceptualizing stress. *American psychologist, 44*(3), 513.
Hofstede, G. (1980). Culture's consequences: International differences in work-related values. In: sage.
Holt, C. C. (1960). Planning Production, Inventories, and Work Force.
Hoppock, R. (1935). Job satisfaction.
Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55.
Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of management journal, 38*(3), 635-672.
Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial economics, 3*(4), 305-360.
Jiang, K., Lepak, D. P., Hu, J., & Baer, J. C. (2012). How does human resource management influence organizational outcomes? A meta-analytic investigation of mediating mechanisms. *Academy of management journal, 55*(6), 1264-1294.
Kahneman, D., Slovic, S. P., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*: Cambridge university press.
Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-292.
Kanter, R. M. (1977). Men and women of the corporation.
Katz, D., & Kahn, R. L. (1966). *Ths social psychology of organizations*: Wiley.
Katz, D., & Kahn, R. L. (1978). *The social psychology of organizations* (Vol. 2): Wiley New York.
Katz, D., Maccoby, N., Gurin, G., & Floor, L. G. (1951). Productivity, supervision and morale among railroad workers.
Katz, D., Maccoby, N., & Morse, N. C. (1950). Productivity, supervision, and morale in an office situation. Part I.
Kendall, M. G. (1948). Rank correlation methods.
Krech, D. (1948). *Theory and problems of social psychology*.
Kuder, G. F. (1946). *Revised manual for the Kuder preference record*: Chicago : Science Research Associates.
Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin, 87*(1), 72.
Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management review, 24*(4), 691-710.
Lawrence, P. R., & Lorsch, J. W. (1967). *Organization and environment managing differentiation and integration*.
Lawshe, C. H., & Alessi, S. L. (1946). Studies in job evaluation: IV. Analysis of another point rating scale for hourly-paid jobs and the adequacy of an abbreviated scale. *Journal of Applied Psychology, 30*(4), 310.
Lawshe Jr, C. (1945). Studies in job evaluation: II. The adequacy of abbreviated point ratings for hourly-paid jobs in three industrial plants. *Journal of Applied Psychology, 29*(3), 177.
Lawshe Jr, C., & Maleski, A. (1946). Studies in job evaluation. 3. An analysis of point ratings for salary paid jobs in an industrial plant. *Journal of Applied Psychology, 30*(2), 117.
Lawshe Jr, C., & Satter, G. (1944). Studies in job evaluation. I. Factor analyses of point ratings for hourly-paid jobs in three industrial plants. *Journal of Applied Psychology, 28*(3), 189.
Leavitt, H. J. (1951). Some effects of certain communication patterns on group performance. *The Journal of Abnormal and Social Psychology, 46*(1), 38.
Lepak, D. P., & Snell, S. A. (1999). The human resource architecture: Toward a theory of human capital allocation and development. *Academy of Management review, 24*(1), 31-48.
Levitt, B., & March, J. G. (1988). Organizational learning. *Annual review of sociology, 14*(1), 319-338.
Likert, R. (1961). New patterns of management.

Likert, R. (1967). The human organization: its management and values.
Lindquist, E. F. (1953). Design and analysis of experiments in psychology and education.
Little, J. D. (1970). Models and managers: The concept of a decision calculus. *Management Science, 16*(8), B-466-B-485.
Litwin, G. H., & Stringer, R. A. J. (1968). *Motivation and organisational climate.*: Division of Research, Harvard Business School, Boston.
Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational behavior and human performance, 3*(2), 157-189.
Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969–1980. *Psychological Bulletin, 90*(1), 125.
MacDuffie, J. P. (1995). Human resource bundles and manufacturing performance: Organizational logic and flexible production systems in the world auto industry. *ilr Review, 48*(2), 197-221.
Maier, N. R. F. (1958). *The appraisal interview: Objectives, methods and skills*. Oxford, England: Wiley.
March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization science, 2*(1), 71-87.
March, J. G., & Simon, H. A. (1958). *Organizations*: John wiley & sons.
Markowitz, H. M. (1959). Portfolio Selection: E cient Diversification of Investments. *Cowles Foundation Monograph, 16*.
Maslow, A., & Frager, R. (1954). Motivation and Personality. In: New York: Harper and Row.
Maslow, A. H. (1943). A theory of human motivation. *Psychological review, 50*(4), 370-396. doi:10.1037/h0054346
Mayfield, E. C. (1964). The Selection Interview - A re-evaluation of published Research. *Personnel Psychology, 17*(3), 239-260. doi:https://doi.org/10.1111/j.1744-6570.1964.tb00065.x
Mayfield, E. C., & Carlson, R. E. (1966). Selection Interview Decisions: First Results From a Long-Term Research Project. *Personnel Psychology, 19*(1), 41-53. doi:https://doi.org/10.1111/j.1744-6570.1966.tb02434.x
McGregor, D., & Cutcher-Gershenfeld, J. (1960). *The human side of enterprise* (Vol. 21): McGraw-Hill New York.
McKinley, J. C., & Hathaway, S. R. (1940). A Multiphasic Personality Schedule (Minnesota): II. A Differential Study of Hypochondriasisa. *The Journal of Psychology, 10*(2), 255-268.
McNemar, Q. (1949). Psychological statistics.
Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American journal of sociology, 83*(2), 340-363.
Miles, M. B., Huberman, A. M., Huberman, M. A., & Huberman, P. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*: SAGE Publications.
Miles, R. E., Snow, C. C., Meyer, A. D., & Coleman Jr, H. J. (1978). Organizational strategy, structure, and process. *Academy of Management review, 3*(3), 546-562.
Mintzberg, H. (1973). The nature of managerial work.
Mintzberg, H. (1978). Patterns in strategy formation. *Management Science, 24*(9), 934-948.
Mintzberg, H. (1983). *Power In and Around Organizations*: Englewood Cliffs, N.J.: Prentice-Hall.
Mintzberg, H., Raisinghani, D., & Theoret, A. (1976). The structure of" unstructured" decision processes. *Administrative Science Quarterly*, 246-275.
Mobley, W. H., Griffeth, R. W., Hand, H. H., & Meglino, B. M. (1979). Review and conceptual analysis of the employee turnover process. *Psychological Bulletin, 86*(3), 493.
Morgan, G. (1986). *Images of Organization*: Sage.
Mowday, R. T., Porter, L. W., & Steers, R. M. (1982). *Employee—organization linkages: The psychology of commitment, absenteeism, and turnover*: Academic press.
Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management review, 23*(2), 242-266.
Nelson, R. R. (1982). An evolutionary theory of economic change. In: harvard university press.
Newcomb, T. M., & Charters Jr, W. (1950). Social psychology.
Nonaka, I., o Nonaka, I., Ikujiro, N., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation* (Vol. 105): OUP USA.
Nunnally, J., & Bernstein, I. (1967). Psychometric Theory, McGraw-Hill Inc. *New York*.

Nunnally, J. C. (1978). *Psychometric Theory: 2d Ed*: McGraw-Hill.
Ocasio, W. (1997). Towards an attention-based view of the firm. *Strategic Management Journal, 18*(S1), 187-206.
Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American psychologist, 17*(11), 776-783. doi:10.1037/h0043424
Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*: University of Illinois press.
Peters, C. C., & Van Voorhis, W. R. (1940). Statistical procedures and their mathematical bases.
Pfeffer, J., & Pfeffer, J. (1981). *Power in organizations* (Vol. 33): Pitman Marshfield, MA.
Pfeffer, J., & Salancik, G. R. (1978). *The external control of organizations: A resource dependence perspective*: Stanford University Press.
Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879.
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. *Annual review of psychology, 63*, 539-569.
Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12*(4), 531-544.
Porter, L. W. (1961). A study of perceived need satisfactions in bottom and middle management jobs. *Journal of Applied Psychology, 45*(1), 1.
Porter, L. W. (1962). Job attitudes in management: I. Perceived deficiencies in need fulfillment as a function of job level. *Journal of Applied Psychology, 46*(6), 375-384. doi:10.1037/h0047808
Porter, L. W. (1963). Job attitudes in management: II. Perceived importance of needs as a function of job level. *Journal of Applied Psychology, 47*(2), 141-148. doi:10.1037/h0041677
Porter, L. W., & Lawler, E. E. (1965). Properties of organization structure in relation to job attitudes and job behavior. *Psychological Bulletin, 64*(1), 23.
Porter, L. W., & Steers, R. M. (1973). Organizational, work, and personal factors in employee turnover and absenteeism. *Psychological Bulletin, 80*(2), 151.
Porter, M. E. (1980). Competitive strategy. *Measuring business excellence*.
Porter, M. E. (1985). Competitive advantage: Creating and sustaining superior performance. In: simon and schuster.
Roethlisberger, F., & Dickson, W. (1939). Management and the worker.
Rokeach, M. (1960). *Open and closed mind*.
Rumelt, R. P. (1974). Strategy, structure, and economic performance.
Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*(5), 529-540. doi:10.1037/0021-9010.62.5.529
Schroder, H. M., Driver, M. J., & Streufert, S. (1967). *Human Information Processing: Individuals and Groups Functioning in Complex Social Situations*: Holt, Rinehart and Winston.
Siegel, S. (1956). Nonparametric statistics for the behavioral sciences.
Smith, P. C., Kendall, L., & Hulin, C. L. (1969). *The measurement of satisfaction in work and retirement.*: Chicago, IL: Rand McNally.
Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47*(2), 149.
Staw, B. M. (1976). Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action. *Organizational behavior and human performance, 16*(1), 27-44.
Staw, B. M., & Fox, F. V. (1977). Escalation: The determinants of commitment to a chosen course of action. *Human relations, 30*(5), 431-450.
Staw, B. M., & Ross, J. (1978). Commitment to a policy decision: A multi-theoretical perspective. *Administrative Science Quarterly*, 40-64.
Stead, W. H., Shartle, C. L., Otis, J. L., Dvorak, B. J., Ward, R. S., Cooper, J. H., . . . Kolbe, L. E. (1940). Occupational counseling techniques: their development and application.
Strong Jr, E. K. (1943). Vocational interests of men and women.

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management review, 20*(3), 571-610.
Super, D. E. (1949). Appraising vocational fitness by means of psychological tests.
Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal, 17*(S2), 27-43.
Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal, 18*(7), 509-533.
Thompson, J. D. (1967). *Organizations in action: Social science bases of administrative theory*: Transaction publishers.
Thorndike, R. L. (1949). Personnel selection; test and measurement techniques.
Thurstone, L. L. (1929). Theory of attitude measurement. *Psychological review, 36*(3), 222-241. doi:10.1037/h0070922
Thurstone, L. L. (1947). Multiple-factor analysis; a development and expansion of The Vectors of Mind.
Tiffin, J. (1947). *Industrial psychology, 2nd ed*. Oxford, England: Prentice-Hall.
Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British journal of management, 14*(3), 207-222.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *SCIENCE, 185*(4157), 1124-1131.
Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *SCIENCE, 211*(4481), 453-458. doi:doi:10.1126/science.7455683
Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, 35-67.
Viteles, M. S. (1953). Motivation and morale in industry.
Viteles, M. S., & Brief, A. (1932). *Industrial psychology*: WW Norton New York.
Vroom, V. H. (1964). Work and motivation.
Wagner, R. (1949). The Employment Interview: a Critical Summary. *Personnel Psychology, 2*(1), 17-46. doi:https://doi.org/10.1111/j.1744-6570.1949.tb01669.x
Webster, E. (1964). *Decision in the employement interview*: Montreal, Eagle Publishing Co., Ltd.
Weick, K. E. (1979). The social psychology of organizing. In: New York : McGraw-Hill, Inc.
Weick, K. E. (1995). *Sensemaking in organizations* (Vol. 3): Sage.
Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal, 5*(2), 171-180.
Whyte. William, H. (1956). Organization Man. In: New York: Doubleday-Anchor.
Wiersema, M. F., & Bantel, K. A. (1992). Top management team demography and corporate strategic change. *Academy of management journal, 35*(1), 91-121.
Williamson, O. E. (1975). Markets and hierarchies: analysis and antitrust implications: a study in the economics of internal organization. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*.
Williamson, O. E. (1985). The economic institutions of capitalism. In: Gabler.
Winer, B. (1971). *Statistical Principles in Experimental Design: 2d Ed*: McGraw-Hill.
Winer, B. J. (1962). Statistical principles in experimental design.
Woodward, J. (1965). Industrial Organization: Theory and Practice. Oxford. In: Clarendon Press.