Improving the Marriage of Modeling and Theory for Accurate Forecasts of Outcomes

Accurately Predicting Precise Outcomes in Business-to-Business Marketing

Arch G. Woodside,

ACCURATELY PREDICTING PRECISE OUTCOMES IN BUSINESS-TO-BUSINESS MARKETING

Arch G. Woodside

ABSTRACT

This chapter identifies research advances in theory and analytics that contribute successfully to the primary need to be filled to achieve scientific legitimacy: configurations that include accurate explanation, description, and prediction – prediction here refers to predicting future outcomes and outcomes of cases in samples separate from the samples of cases used to construct models. The MAJOR PARADOX: can the researcher construct models that achieve accurate prediction of outcomes for individual cases that also are generalizable across all the cases in the sample? This chapter presents a way forward for solving the major paradox. The solution here includes philosophical, theoretical, and operational shifts away from variable-based modeling and null hypothesis statistical testing (NHST) to case-based modeling and somewhat precise outcome testing (SPOT). These shifts are now occurring in the scholarly business-to-business literature.

Keywords: Accuracy; industrial marketing; management; modeling; prediction; science

Improving the Marriage of Modeling and Theory for Accurate Forecasts of Outcomes
Advances in Business Marketing & Purchasing, Volume 25, 63–84

ISSN: 1069-0964/doi:10.1108/S1069-096420180000025006

INTRODUCTION: ACHIEVING SCIENTIFIC LEGITIMACY

LaPlaca and colleagues (Hadjikhani & LaPlaca, 2013; LaPlaca, 1997; LaPlaca & da Silva, 2016) described in-depth the first paradigm shift in business-to-business (B-to-B) research from description and explanation of business exchanges based on transactions to description and explanation of business exchanges based on relationships. Equally important, they identify what is still necessary to accomplish for B-to-B research to achieve scientific legitimacy, "B2B relationships as a subject of scientific enquiry will need to seriously engage into what can be termed a true paradigm shift, one that advances discovery in this area from sheer descriptive analysis and reporting to the development of explanatory schemata and theoretical frameworks of a kind that allow for more accurate prediction of underlying B2B phenomena" (LaPlaca & da Silva, 2016, p. 232).

LaPlaca and colleagues provide foundation insights into the steps necessary to be taken to achieve scientific legitimacy, including embracing prediction and control as necessary objectives in B-to-B research – research focusing on description and explanation is necessary but insufficient for advancing science in the B-to-B discipline. "In conducting scientific investigations, researchers, particularly scientists studying physical phenomena, progress through a hierarchy of types of research: descriptive, explanatory, predictive, and control (LaPlaca, 2013). The ultimate goal of science is to control events where possible … Improved understanding and predictive capabilities will reduce marketing errors and improve overall marketing effectiveness and efficiency. In this way, B-to-B marketing research will truly make a contribution to society" (LaPlaca & da Silva, 2016, p. 232).

The following discussion focuses on how to accomplish the true paradigm shift that LaPlaca and colleagues identify. The study here provides examples of research contributing to knowledge and theory that advance prediction and control in B-to-B contexts. The study indicates that shifting beyond linear model construction and symmetric tests (i.e., multiple regression analysis (MRA) and structural equation modeling (SEM)) and embracing complexity theory and asymmetric tests (i.e., constructing and testing algorithms by "computing with words," Zadeh, 1996, 2010) are necessary steps to be taken to accomplish the true paradigm shift. Researchers in B-to-B research benefit from recognizing that the current dominant logic of performing null hypothesis testing (NHST via MRA and SEM) is "corrupt research" (Hubbard, 2016) practice and from recognizing that predicting by algorithms via somewhat precise outcome testing (SPOT) advances B-to-B research toward achieving scientific legitimacy.

Following this introduction, the second section answers the question, "Predicting what – directions or outcomes?" The third section provides examples of predicting precise outcomes in the B-to-B research literature. The fourth section expands on prior calls (Misangyi et al., 2017; Woodside, 2014) to embrace complexity theory as the foundational philosophy in B-to-B research – the expansion includes a description of "four-corner modeling" via predictive algorithms of complex (versus the currently dominant single condition) outcomes. The fifth section concludes this tribute by elaborating on how to overcome naysayers and "the forces of inertia" (Huff, Huff, & Barr, 2001) that usually serve to prevent adoption of superior theory and method. It also addresses the question, what steps are helpful for overcoming these forces to gain acceptance of research using SPOT rather than NHST by reviewers and editors in scholarly journals? The essay here and conclusion support the conclusion that the teen-years of the 21st century bear witness to B-to-B researchers' successful responses – finally – to LaPlaca's call for a truly new paradigm shift.

PREDICTING WHAT – DIRECTIONS OR OUTCOMES?

Along with convincingly demonstrating that the significance difference paradigm is methodologically impaired and statistically broken and "embedded in an academic social structure whose publication biases complete the institutionalizing this corruption" (Hubbard, 2016, p. 9), he raised the point that "there is no reason why theories in the management social sciences cannot yield precise (or interval) predictions…this line of thinking flies in the face of conventional wisdom that theories in these areas are unable to specify point predictions" (Hubbard, 2016, pp. 192–193). In his demonstration of the null value of NHST, Hubbard (2016) reviews more than 50 studies that are consistent with Schmidt's (1996, p. 116) conclusion: "We must abandon the statistical significance test." Trivial findings include findings that a difference between two means is not zero, partial regression weights for variables in a regression model are not equal to zero (cf. Cohen, 1994, p. 1000), or two variables have a positive or a negative relationship. "Thus asking, 'Are the effects different?' is foolish." What we should be answering first is, "Can we tell the direction in which the effects of A differ from the effects of B?" (Tukey, 1991, p. 1000). However, what Tukey (1991) proposed turns out to be foolish as well. The better, more informative questions to ask and answer include, "Within what complex conditions does high A indicate high B, low A indicate high B, low A indicate high B, and low A indicate low B." If both are continuous variables, converting each to quintiles and cross-tabulating the two sets of cases almost always demonstrates that cases occur in all 25 cells. Even when a main effect is large indicating "A" associates with "B," cases found to be in the cells indicating associations contrary to the main effect are not merely unexplainable blips – such "seeming anomalies" are deserving of explanation and predictive modeling.

Directional findings (e.g., $r = .57$, $p < .01$) are qualitative predictions offering scant substantive information leading McCloskey (2002, p. 55) to describe almost all the harm such studies inflict on the discipline – what she labels the "Two Sins of Economics" (i.e., being content with only qualitative predictions in both theory and applied work):

> The progress of economic science has been seriously damaged. You can't believe anything that comes out of the Two Sins. Not a word. It is all nonsense, which future generations of economists are going to have to do all over again. Most of what appears in the best journals is unscientific rubbish. I find this unspeakable sad. (McCloskey, 2002, p. 55)

Directional testing and tests of significance differences are bad science for additional reasons. As practiced in articles in the best journals, they fail to indicate when exceptions occur to the directions supported by the statistical tests. Given that in real-life exceptions almost always occur to a statistically significant main effect, modeling the causes leading to the contrarian directional outcomes would likely provide important findings. The current practice in the dominant paradigm of testing the relative size of influence of independent variables in linear regression and SEM research represents a mismatch between theory and analytics (Fiss, 2007). Variables' weights in MRA/SEM are competing with one another for indicating that each variable has a significant positive or negative influence in these models. If the associations among two independent variables are both large between them and with the dependent variable, one of the two appears to be non-significant in the resulting model due to this "multicollinearity."

In human resources research attempting to construct models predicting highly competent managers (managers in the top quintile of competence), McClelland's (1998) frustration with the severely limited usefulness of regression findings, and his decades of experience and insights in working in data analysis, led him to try discretizing variable data into quintiles and creating algorithms. Thus, McClelland (1998) shifted his theory construction and analytics from variable-based to case-based reasoning. McClelland (1998) was able to construct somewhat precise outcome tests (SPOT) ("SPOT" is not a term used by McClelland) that were highly accurate in identifying highly competent managers among samples of managers not used in the construction of the models (i.e., the algorithmic model had high predictive validity). While McClelland's (1998) work has had high impact (1000+ citations by 2017), his method has been widely ignored. When SPOT findings are "useful" (avoiding "statistically significant" here), all or nearly all cases having high scores in the asymmetric model have high scores in the outcome. For example, cases (managers) with high scores across all causal conditions in McClelland's antecedent conditions were identified to be highly competent managers. McClelland's (1998) hit (accuracy) ratios for identifying highly competent managers were frequently above 7-to-1. McClelland's analytics are an example of statistical sameness tests of precise outcomes – a case-based approach to data analysis – rather than using NHST. Hubbard (2016, p. 5) points out, "Looking for reproducible results is a search for significant sameness, in contrast to the emphasis on the significant difference form a single experiment" (Nelder, 1986, p. 113).

The great power in using MRA and SEM to generate models having high fit validity cannot be denied. In fact, because these analytics make use of all the information available in the data, highly significant terms ("paths") in these models occur even when using a table of random numbers for data (Armstrong, 2012). But the proof is in testing for predictive validity of models by seeing how well they predict outcomes for cases in separate samples from the cases used to create the models. "Achieving a good fit to observations does not necessarily mean we have found a good model, and choosing the model with the best fit is likely to result in poor predictions. Despite this, Roberts and Pashler (2000) estimated that, in psychology alone, the number of articles relying on a good fit as the only indication of a good model runs into the thousands" (Gigerenzer & Brighton, 2009). These studies are examples of shallow analysis that are accurately describable as examples of the rubbish that saddens McCloskey (2002).

The pervasive practice of researchers using NHST is to universally fail to examine the occurrence of reversals in relationships that occur almost always in data sets for 10–20% of the cases in a data set – even when the effect size is large ($r^2 \geq 0.25$) for a relationship. Complexity theory (Urry, 2005; Wu, Yeh, Huan, & Woodside, 2014) indicates the occurrence of such cases (e.g., X decreases indicate Y increases) even though the main relationship is that X increases associate with Y increases. Such contrarian cases are verifiable easily by creating quintiles for both X and Y variables and cross-tabulating the quintiles (Woodside, 2016).

PREDICTING PRECISE OUTCOMES IN THE B-TO-B LITERATURE

A few studies are identifiable in the literature that include the use of SPOT and predictive validation of the predictions. These studies are illustrative of several good science principles. For example, these studies construct asymmetric causal models – that is, they recognize that causal configurations indicating a negative outcome are not the mirror opposite of the causal configurations indicating a positive outcome. The vast majority of contexts in all management and behavioral sciences present asymmetric and not symmetric cause-outcome associations. Consequently, the use of symmetric tests (e.g., *F*-tests, *r*, MRA, and SEM) has limited and shallow usefulness for understanding, explanation, and prediction of positive and negative outcomes. This section briefly discusses four studies that include SPOT, asymmetric models, with predictive validation.

Describing, Explaining, and Predicting Specific Price-Point Shifts

The study of wholesale price changes by Howard and Morgenroth (1968) (hereafter H&M, 1968) and Morgenroth (1964) is an early example of asymmetric modeling with testing for predictive validity. Both articles report in-depth on the same study. The study included a triangulation of data collection: multiple face-to-face interviews of the same persons participating in pricing decisions in a specific context, direct observation, document analysis, and confirmatory interviews with additional executives in the same firm who were located in other sales territories. Fig. 1 is a summary of the model.

In Fig. 1, the descriptive decision model derived from the research appears in flow chart form in Fig. 1 with price as its output. Fig. 1 depicts a sequential binary choice process. The cue or triggering element of the process is Box 1 in the upper left-hand corner of Fig. 1, labeled "Watch Pwilt," which means, "Watch the wholesale (w) price (P), of the initiator (i) in each local (1) market at each point in time (t)" (H&M, 1968, p. 420). A verbal and Boolean algebra explanation of parts of the model appears on the right-hand side of Fig. 1. In total, findings in Fig. 1 indicate that price increase decisions are easier, quicker, and include fewer participants usually than price decrease decisions.

The predictions of the model were tested in two ways. Filing cabinets (no computers at the time of the study) contained pricing data and decisions of the division over a six-year period. A systematic sample of every tenth filing was taken. The filings were arranged internally in chronological order, with the date that a competitor's move was initially made (the triggering) serving as the specific criterion of order. This sample yielded 32 decisions which were compared with the decisions predicted by the model. A comparison of predictions and observations is made in the study. In addition, 130 other decisions in other divisions of the company were used to test the model (Morgenroth, 1964, p. 21). Fig. 3 appearing in both Morgenroth (1964) and in H&M (1968) reports perfect agreement for the predicted and observed outputs. "Hence, the hypothesis that the model represents the executive's decision process is confirmed by the output test" (H&M, 1968, p. 424).

Describing, Explaining, and Predicting Supermarket Buying Committee Decisions to Accept/Reject Manufacturers' New Brands and Brand Extensions

Montgomery (1975) includes both symmetric and asymmetric models in his study of food stores' buying decisions to accept versus reject new brands' and brand extensions' offerings by manufacturers. Montgomery (1975, p. 256) describes the study, "Interviews with buyers and attendance at buying committee meetings were used to identify a list of potentially important variables in the decision to accept or reject a new product for distribution in a supermarket.

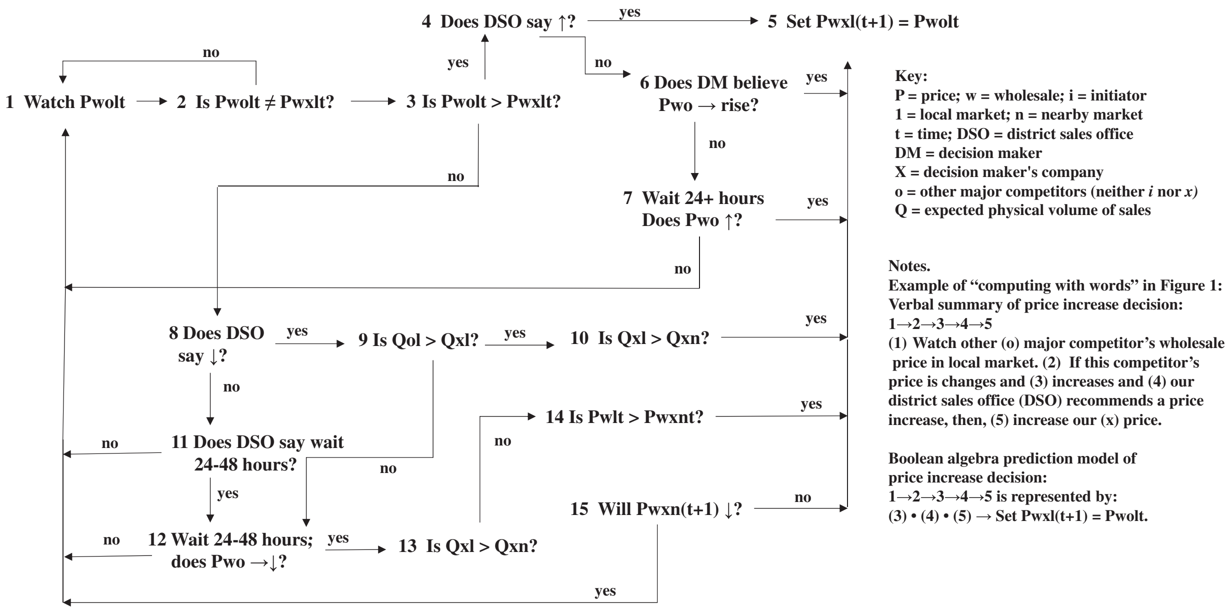


Fig. 1. Flow Diagram of a Pricing Decision Prediction Model. *Source*: Adapted from Morgenroth (1964, p. 19).

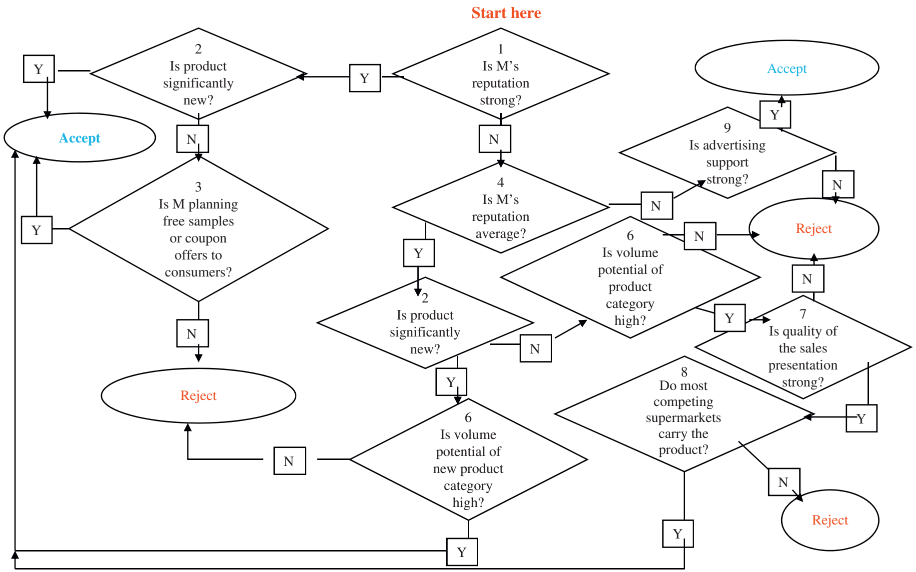
This list was then used to develop a structured personal interview. The interview consisted of a buyer rating a proposed new product on each of the variables. In addition, it was ascertained whether the product was accepted or rejected. The accept/reject result and variable ratings were obtained by personal interview for 124 products proposed to 3 supermarket buyers [each for a separate supermarket store chain] in the Boston area." Montgomery (1975) reviews a series of problems with the use of symmetric analysis which led him to perform an asymmetric modeling exercise – what he refers to as a "gatekeeper analysis." His criticisms of symmetric testing and his proposals for asymmetric (SPOT) testing were about 30+ years ahead of similar arguments in the management and marketing literatures (e.g., Fiss, 2007; Woodside, 2013).

Fig. 2 is a summary of Montgomery's asymmetric models. Model 1 is the simplest model: R•N→Accept. That is, if the buying committee judges the manufacturer's reputation to be "strong" and the product to be very new (i.e., top quintile in "newness"), then the product is accepted as a new SKU (stock-keeping unit). No additional information cues are used by the committee; Model 1 is asymmetric and does not include the attempt to predict rejection. Note that the models in Fig. 2 fulfill the same tenets as the models in Fig. 1: equifinality, causal asymmetry, and no one condition being necessary or sufficient for accept or reject outcomes.

This "gatekeeping" analysis will produce a tree diagram which is similar to those produced by automatic interaction detection (AID). A prime contrast between AID and the proposed procedure is that AID seeks to maximize between-group sums of squares relative to within-group sums of squares while the gatekeeping procedure looks for points of near non-overlap in the sample distributions. One advantage of the gatekeeping analysis is that it may be used on relatively small samples. In the case presented here there are insufficient data to run an AID analysis, whereas the gatekeeper analysis is feasible.

Overall, the gatekeeping analysis provides nearly 93% correct classification in contrast to 86% for discriminant analysis. It correctly classifies 95% of all rejected products and 86% of all accepted products in contrast to the 84% of rejects and 92% of accepts from discriminant analysis. Thus, the gate-keeping analysis performs relatively better on the rejected products and relatively poorer on the accepted products.

Montgomery (1975, p. 263) contrasts the gatekeeper and the linear discriminant analyses, "The gate-keeper analysis makes no assumptions as to the scale of the data nor its underlying distribution. It also allows for nonlinear conditional interaction among the variables. Further, the gatekeeper approach would seem to be a closer approximation to a buyer's decision process. It is difficult to imagine that a buyer cerebrally forms a weighted linear combination of variables and compares this score to a cut-off level as is done in linear discriminant analysis. It seems more plausible that they exhibit a decision process something like the gatekeeper tree where serious failure at some point spells difficulty for a product offering. The gatekeeper tree is also a useful data summarizer for



PathBoolean Expression, Key

1. 1-2R•N → A, R = Reputation; N = New; A = Accept
2. 1-2-3 (a)R•~N•F → A, F= Free samples
3. 1-2-3 (b)R•~N•~F → R, ~ = Not; R = Reject
4. 1-4-9~R•~RA•S → A, RA= Reputation average; S = ad support
5. 1-4-9 (b)~R•~RA•~S → R
6. 1-4-2-6 (a)~R•RA•N•V → AV = Volume potential
7. 1-4-9-6 (b)~R•RA•N•~V → R
8. 1-4-2-6 (b)~R•RA•~N•~V → R
9. 1-4-2-6-7~R•RA•~N•V•~Q → R
10. 1-4-2-6-7-8~R•RA•~N•V•Q•C → A, Q = quality of sales presentation C = Competitors' carry new product
11. 1-4-2-6-7-8 (a) ~R•RA•~N•V•Q•C → A
12. 1-4-2-6-7-8 (b) ~R•RA•~N•V•Q•~C → R

Notes. Mid-level dot ("•") indicates the conjunctive "and". The horizontal arrow ("→") points to an accept or reject outcome.

Fig. 2. An Ethnographic Decision Process Model of Supermarket Committee Buying Decisions about a Manufacturer's (M's) New Product Offering. *Source*: Adapted from Montgomery (1975) with new expressions of causal mechanism.

communication with management. It is simple and quickly understood. Most market researchers, to say nothing of management, have difficulty in fathoming the intricacies of discriminant analysis. Finally, from an empirical point of view it fits the data better, 93% vs. 86% correct classifications."

Montgomery (1975, p. 263) provides a partial attempt of providing predictive validation, "Since responses from three buyers were pooled in this sample, the error rates for the individual buyers should be examined. For the first buyer the gatekeeper analysis correctly classified 62 out of 66 products or 94%. For the second, 28 of 30 products or 93.3% were correctly classified. For the third buyer, 25 of 28 products or 89.3% were correctly classified. Hence, the results are quite consistent across buyers. The process model diagram in the figure remains, of course, a composite of all three buyers." A better approach would have been to construct separate models for each buyer and report the hit/miss ratios for new product acceptance and rejection across the three buyers as well as the similarities and any differences in the accept and reject models among the three buyers. The results of such testing might indicate that one model among the three buyers was more generalizable than one or both models for the other two buyers. Possibly such an analysis might still be possible since the data came from a master's thesis at MIT (Phillips, 1968) and might be found in the thesis.

Ethnographic Decision Tree Models (EDTM) and Field Experiments of Farmers' Adoption Choices

The research by Gladwin (1989a, 1989b) and associates (Cough, Gladwin, & Hildebrand, 2002) includes constructing asymmetric ethnographic models of African women farmers' decisions to adopt or reject agricultural-related innovations. This body of work includes testing the predictive validity of the models and field experiments (the control dimension in achieving B-to-B scientific legitimacy). The field experiment examined alternative "starter packs" to all farming households in Malawi. The starter packs contained small packs of hybrid maize seed, fertilizer, and either groundnuts or soybeans. The 1999 starter pack distribution also included a pilot voucher project that distributed two different types of vouchers, in a test to see whether the vouchers received by some of the farmers were more effective than the packs received by other farmers. The purpose of this study was to evaluate that test. Gough et al. (2002) examined the differences between the three distribution systems of the starter pack, starter-pack voucher, and flexi voucher to estimate which was the more effective tool for improving food security among Malawian smallholder farmers.

Gough et al. (2002) report that the results showed the most economically enhancing tool for smallholders, especially the poorest, was flexi vouchers. The benefit of distributing flexi vouchers was manifested through increased household cash income (averaging MK (Malawian Kwacha, the currency of Malawi) 2143 in group-four) and not maize production. Cough et al. (2002) concluded this MK 2143 would purchase 450 kg of maize over the five-year period, enough to feed a chronically food-insecure family (with food requirements of 700 kg of maize per year) for 7.7 months spread over the five-year time period, or 1.54 months per year. Unfortunately, this additional maize is probably not enough to make a chronically food-insecure household in Malawi food-secure. It is just too little in the Malawi situation where food-insecure households now face hunger seasons of five to six months.

The work by Cough et al. (2002) is part of the behavioral science domain of "developmental economics." The field of developmental economics focuses in particular on the issue of control – what programs can be designed and implemented by government and non-government organizations (NGOs) to improve the quality of life in a community and/or nation? Several developmental economists report field experiments with random assignment of cases to various treatment and control groups (Banerjee & Duflo, 2008). Most of the work in developmental economics relies solely on symmetric tests of outcomes. The shift to asymmetric testing is yet to occur with few exceptions (Cough et al., 2002 is an example of exception). However, several studies in developmental economics are examples of controlled experiments in B-to-B contexts (e.g., Duflo, Kremer, & Robinson, 2008) that by their very nature focus on measuring the effectiveness of alternative control treatments.

Additional control-centered research directly in B-to-B context supports the perspective that such research contributes to both theory and practical usefulness. This work includes Wilson's (2010) three field experiments, focusing on incorporating ways to discourage shopping-cart abandonment by B-to-B customers and the use of two different free-shipping promotions, which were used as the basic research methods for collecting the data. Web traffic conversion funnels are used to conduct the analysis and present the findings. His findings support the conclusion that clickstream data using web analytics procedures serves as a useful tool in the enhancement of a B2B website by investigating how visitors move through the website conversion process and complete their purchase. Improved sales result from each of the three field experiments. Tucker and Zhang (2010) use field experiment data from a B-to-B website to examine the efficacy of these different display formats. Before each potential seller posted a listing, the website randomized whether to display the number of buyers and/or sellers, and if so, how many buyers and/or sellers to claim. Tucker and Zhang (2010) find that when information about both buyers and sellers is displayed, a large number of seller deters further seller listings. "However, deterrence effect disappears when only the number of sellers is presented. Similarly, a large number of buyers is more likely to attract new listings when it is displayed together with the number of sellers. These results suggest the presence of indirect network externalities, whereby a seller prefers markets with many other sellers because they help attract more buyers" (Tucker & Zhang, 2010, p. 805). While the study by Tucker and Zhang (2010) relies on conventional theory symmetric tests (regression analysis), Woodside, Schpektor, and Xia (2013) describe how to test for statistical sameness outcomes (SPOT) with data from field experiments.

Asymmetric Modeling in Industrial Marketing Management

Brenes, Ciravegna, and Woodside (2017) describe asymmetric models of implemented strategy and competitive advantage for ROE, negation of ROE, and complex outcome statements for agribusiness firms ($n = 247$) across seven Latin America nations as well as tests the predictive validities of models across specific nations for the models of sampled firms within Costa Rica, El Salvador, Guatemala, and Nicaragua. The findings support the proposition that constructing complex antecedent statements (i.e., algorithms/configurations/recipes/screens) is useful for indicating high performance or the negation of high performance consistently. Configural implemented strategy models have direct influences on both high and low performance outcomes, while competitive advantage models impact low, but not, high performance outcomes. Complex competitive advantage conditions contribute indirectly to high performance outcomes.

Brenes et al. (2017) findings support the perspective that competitive orientation and product portfolio planning tools are shallow and misleading approaches to the advancing useful strategy theory (cf. Anterasian, Graham, & Money, 1996; Armstrong & Collopy, 1996; Armstrong & Green, 2007). "Embracing the core theoretical tenets of complexity theory is necessary for theory to respond and to adequately answer the crucial problem in strategy theory (Powell, Lovallo, & Fox, 2011) − accounting for firm heterogeneity. Complexity theory tenets coupled with asymmetric modeling using Boolean algebra focus on identifying outcomes of interest (e.g., high ROE) consistently. This approach provides for parsimonious but not overly simplistic solutions that occur from building models to explain the relative importance of terms in regression models via symmetric tests using matrix algebra. As Fiss (2007) explains and demonstrates (Fiss, 2011) we can overcome the mismatch that now dominates strategic theory by matching case-based theory with case-based analytics" (Brenes et al., 2017, pp. 17−35).

EMBRACING COMPLEXITY THEORY AS THE FOUNDATIONAL PHILOSOPHY IN B-TO-B RESEARCH

In marketing, Kotler (1967, p. 1), famously pronounced, "Marketing decisions must be made in the context of insufficient information about processes that are dynamic, nonlinear, lagged, stochastic, interactive, and downright difficult." Yet the substantial majority of studies in the nearly five decades since this pronouncement continue to ignore all the decision features that Kotler describes. Gummesson (2006, 2008) urges marketing scholars and educators to accept the complexity of marketing and develop a network-based stakeholder approach – balanced centricity – epitomized by the concept of many-to-many marketing. Gummesson (2008) calls for a rejuvenation of marketing. "Reality is complex whether we like it or not. This is where network theory comes in. Its basics are simple; a network is made up of nodes (such as people or organizations) and relationships and interaction between those. Network theory is part of "complexity theory," recognizing that numerous variables interact, that the number of unique situation is unlimited, that change is a natural state of affairs, and that processes are iterative rather than linear … But is balanced centricity a realistic objective or is it yet another professorial whim? I do not have the answer but I am convinced that if we keep fragmenting marketing and other business functions and duck complexity, context, and dynamics, we will not move ahead. A change requires that we reconsider marketing basics and abandon mainstream methodological rigidity and move toward a more pragmatic and holistic research agenda" (Gummesson, 2008, pp. 16, 17).

Scholars before Gummesson (2008) describe the need to reconsider mainstream methodological rigidity and move toward more pragmatic and holistic (i.e., patterns or systems) research agenda. Bass, Tigert, and Lonsdale (1968) offer evidence that the contention that the low R^2s obtained in regression analysis lead to false conclusions about the ability of socioeconomic variables as well as attitudinal measures to substantially explain variance in dependent variables since R^2 is a measure of a model's ability to predict individual rather than group behavior. McClelland (1998) goes further in stressing that most researchers do not really want to explain variance in dependent variables; what they want to do is to describe, explain, and accurately predict high scores in an outcome condition (i.e., create algorithms – decision rules – that work almost all the time in providing an effective decision and avoiding bad decisions). Without likely being aware of McClelland's (1998) contributions to asymmetric thinking, research methods, and parsimony, Ragin (2000, 2006, 2008) relies on Boolean algebra rather than the dominating use of matrix algebra-based statistical methods to offer parallel insights and methods in sociological research and beyond.

Three additional points need stressing that relate to complexity theory's focus on patterns in phenomena. First, "Scientists' tools are not neutral" (Gigerenzer, 1991). Research methods and instruments shape the way we think and test theories. Thus, reviewers' question whether a given paper is trying to make a contribution to theory or method sometimes misses the point that a research paper tries to do both – as is the case here. Second, reports of model confirmation relying only on fit validity need to stop; reports that partial regression coefficients in an MRA model are significant are insufficient findings and of limited usefulness. Analysts assume that models with a better fit provide more accurate forecasts. This view ignores the extensive research showing that fit bears little relationship to ex ante forecast accuracy, especially for time series. Typically, fit improves as complexity increases, while ex ante forecast accuracy decreases as complexity increases, a conclusion that Zellner (2001) traces back to Sir Harold Jeffreys in the 1930s (Armstrong, 2012). Gigerenzer and Brighton (2009) provide substantial empirical evidence supporting the focus for accuracy and theory advancement via predictive validity and not fit validity. Third, "Developing the full potential of complexity theory, especially in the social sciences, requires more rigorous theory development and fewer popular articles extolling the virtues of the 'new paradigm', more studies testing the new theories and fewer anecdotal claims of efficacy, greater development of tools tailored for particular contexts, and fewer claims of universality. Without such rigor, social scientists face the danger that, despite its high potential, 'complexity theory' will soon be discarded, perhaps prematurely, as yet another unfortunate case of physics envy" (Sterman & Wittenberg, 1999, p. 338).

Complexity Theory Tenets

The following tenets (Ti) are steps to contribute rigor in response to Sterman and Wittenberg's (1999) call to do so. T.1: A simple antecedent condition may be necessary, but a simple antecedent condition is rarely sufficient for predicting a high or low score in an outcome condition. T.2: A complex antecedent condition of two or more simple conditions is sufficient for a consistently high score in an outcome condition – the recipe principle. T.3: A model that is sufficient is not necessary for an outcome having a high score to occur – the equifinality principle. T.4: Recipes indicating a second outcome (e.g., rejection) are unique and not the mirror opposites of recipes of a different outcome (e.g., acceptance) – the causal asymmetry principle. T.5: An individual feature (attribute or action) in a recipe can contribute both positively and negatively to a specific outcome depending on the presence or absence of the other ingredients in the recipes. T.6: For high Y scores, a given useful recipe (i.e., model) is relevant for most but not all cases; coverage is less than 1.00 for any one recipe (e.g., a specific useful model may be accurate in predicting high outcome scores for the majority (7 of 8, 14 of 15, 25 of 27) cases but a few false positives occur – thus, the expression, "somewhat precise outcome testing"). T.7: Exceptions occur for high X scores for a given recipe that works well for predicting high Y scores. T.8: Discretizing continuous variables using quintiles and cross-tabulating frequently identify 10–20% of the cases to be contrary to a medium-to-large symmetric main effect; consequently, modeling the four corners of configural two cross-tabbed conditions will deepen description, explanation, and predictive knowledge in B-to-B research.

The Proposal for Four-Corner Modeling

Four-corner modeling includes constructing causal configurations of complex antecedent statements for cases in each set of cases in the four corners of cross-tabs of quintiles or deciles of cases for two or more variables. With sample sizes reasonably large (e.g., $n > 50$), discretizing (particularly useful to do by quintiles) two continuous variables and cross-tabulating the two sets of cases result in a few-to-many cases appearing in all 25 cells – if using quintiles – even when the main effect size between the two variables is large ($r^2 \geq 0.25$). A vast majority of studies using the current symmetric-testing dominant logic ignore the occurrences of cases that show associations contrary to statistically significant positive or negative directional associations.

Fig. 3 in this chapter includes a large positive main effect finding in a study of core self-evaluations (CSEs) and job satisfaction for 1,000 New Zealand farm managers (Ang & Woodside, 2017). Judge, Locke, and Durham (1997) propose the following four CSEs as indicators of a higher order construct, a positive self-concept: self-esteem, generalized self-efficacy, locus of control, and emotional stability (low neuroticism). Unlike prior work focusing on establishing that each of the CSEs has a positive significant correlation with job satisfaction (JS) and that combining the four traits to form a single latent construct (i.e., positive self-concept) associates positively with JS, Ang and Woodside (2017) proposed, tested, and confirmed that cases occur where low scores on some of the traits occur for cases (individuals) having high JS, and high scores on some of the traits occur among cases having low JS. Such cases are more than unexplainable blips. Such cases likely are due to contextual influences that are accountable by using asymmetric rather than symmetric modeling. Ang and Woodside's (2017) study contributes by proposing and illustrating a paradigm shift from variable-based theory construction and symmetric testing to case-based theory construction and asymmetric testing.

Additional studies are now becoming available that include the use of four-corner analysis (Feurer, Baumbach, & Woodside, 2016; Nagy et al. 2017). For example, rather than just constructing a directional market of how market conditions affect firm performance, four-corner modeling has been useful for constructing causal models that indicate firms performing well in declining markets and additional models indicating firms performing poorly in growing markets (Nagy et al., 2017). As McClelland (1998) indicated, converting continuous variables into quintiles is more than just a data manipulation exercise; such conversions represent a shift from variable-focused to case-focused research. A researcher can expect to find a combined 10–20% of the cases that display contrarian associations to hypothesized and confirmed directional associations. Reporting only supported directional hypotheses is shallow research because of the presence of such contrarian cases that are anomalies to the directional hypotheses. An anomaly is a fact or case that does not fit received wisdom. "To

Job Satisfaction

CSE Group	Very low	Low	Middle	High	Very High	Total
Very low	85	53	39	20	8	205
Low	32	49	74	28	20	203
Middle	16	42	74	33	38	203
High	13	22	78	50	53	216
Very high	2	5	30	27	109	173
Total	148	171	295	158	228	1,000

C1 C2 C3 C4

Phi = 0.63, $p < .001$

= the number in the box indicates the most frequent number of cases in the row.

= the number in the dotted-line boundary are cases contrary to the highly significant statistically positive linear relationship indicated by phi = 0.63; the contrarian cases have very low and low CSE scores but very high and high in job satisfaction or cases having high and very high CSE scores but low or very low job satisfaction scores.

Four-corners' shares of cases:

Supporting the positive main effect:

Corner 1: Low S AND low CSE: 85+53+32+49 = 219 (21.9%)

Corner 4: High S AND high CSE: 109+27+53+50 = 239 (23.9%)

Contrary to the positive main effect:

Corner 2 (expecting low but high S): 20+8+28+20 = 76 (7.6%)

Corner 3 (expecting high but low S): 13+22+2+5 = 42 (4.2%)

Four-corner analysis includes constructing/testing complex outcomes–four sets of models to answer the following questions:

- What causal models predict CSE•S cases consistently?
- What causal models predict ~CSE•~S cases consistently?
- What causal models predict ~CSE•S cases consistently?
- What causal models predict CSE•~S cases consistently?

Key: "~" = negation; "•" = logical "AND" condition.

Fig. 3. Cross-Tabulation of Quintiles of Cases for Core Self-Evaluations (Summed CSE Averages) and Job Satisfaction. *Source*: Adapted from Ang and Woodside (2017, p. 36).

a certain kind of mind, an anomaly is an annoying blemish on the perfect skin of explanation. But to others, an anomaly marks an opportunity to learn something perhaps very valuable. In science, anomalies are the frontier, where the action is" (Rumelt, 2011, pp. 247−248). Shifting from the current dominant variable-based logic to case-based logic increases the possibilities of describing, explaining, and predicting cases having anomalous properties.

CONCLUDING REMARKS

Fig. 4 is a visual summary of core features in the current dominant logic in B-to-B research and core features of a true new paradigm now in the introduction stage in the discipline. The features appear in two Venn diagrams in Fig. 4 to represent the configurational nature of research paradigms − rather than a house-of-cards, research paradigms include self-reinforcing joined-at-the-hip forces. Replacing one bad feature alone is insufficient. All features need replacing or dramatically improving by a truly new research paradigm. The evolutionary rise in the current dominant variable-based mostly description- and explanation-focused logic in B-to-B research occurred in the 1950s and continued to the end of the 20th century. The revolutionary introduction of a true, new, case-based paradigm focusing mostly on description, explanation, and prescription is occurring in the second decade of the 21st century. Growth is expanding rapidly now (2015−2019) in the number of scholarly articles featuring the true new paradigm (Roig-Tierno, Gonzalez-Cruz, & Llopis-Martinez, 2017).

Twenty paradigm shift-catalysts appear in the center of Fig. 4. These shift-catalysts are essays and mostly non-NHST SPOT-empirical studies that include features and full-blown expositions of a true new research paradigm. The 20 catalysts include Hubbard's (2016) thorough documentation of the corrupt practices of NHST − the foundational analytical stance of the current dominant logic. Because NHST is a bad science practice, the editor of one prestigious scholarly journal (*Basic and Applied Social Psychology*) announced that authors of all future articles accepted for publication would need to remove reports of statistical significance tests before their articles were published (Trafimow & Marks, 2015). NHST is more than a tool for data analysis; the use of NHST suggests embracing a theoretical stance. Unfortunately, the current dominant logic and use of correlations, *F*-tests, MRA, and SEM nurture the perspective that NHST is the only scientific testing procedure worthy of using. Reading Hubbard (2016) is very helpful for overcoming such a sad and wrong conclusion. Woodside (2017) expands on Hubbard's (2016) call to use "statistical sameness" outcome testing by presenting several studies that do just that.

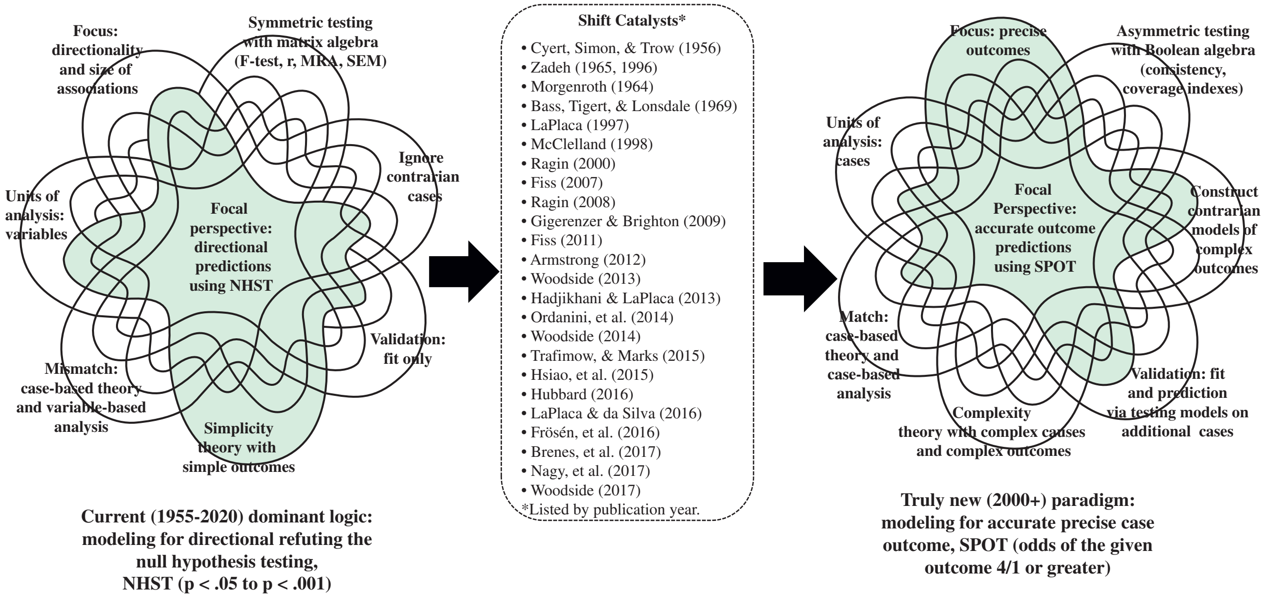


Fig. 4. Manifesto for a True Paradigm Shift in Business-to-Business Research.
Acronyms: *F*-test = analysis of variance; *r* = correlation; MRA = multiple regression analysis; SEM = structural equation modeling; NHST = null hypothesis statistical test; SPOT = somewhat precise outcome test.

Cyert, Simon, and Trow (CST, 1956) is the first shift-catalyst appearing in Fig. 4. This study is a foundational (iconic) contribution in the B-to-B discipline specifically and the field of decision sciences in general. The authors' nitty-gritty focus on "observation of a business decision" supports the proposition that researchers need to keep their focus firmly fixed on case-based theory construction and testing of complex prescriptions of outcomes within given contexts. Some of the shift-catalysts in Fig. 4 illustrate the usefulness (sometimes, the necessity) of including both SPOT and NHST findings in the same study as a procedure to gain acceptance from the forces of inertia and naysayers for such radically new features in the true new paradigm. Journal article reviewers usually perceive new tools, new theories, and unexpected findings to be controversial and they frequently reject such studies (Armstrong & Green, 2007). "Extensive evidence on peer review shows that papers with findings that contradict important viewpoints are nearly always rejected by reviewers (Armstrong, 2009). For example, a survey by Armstrong and Hubbard (1991) found that: 'Editors of 16 psychology journals reported that reviewers dealt harshly with papers that contained controversial findings'" (Armstrong & Green, 2007, p. 123). Armstrong found that none of what he considers his 20 most important papers received full acceptance by reviewers (Armstrong & Green, 2007). Ordanini, Parasuraman, and Rubera (2014) and Frösén, Luoma, Jaakkola, Tikkanen, and Aspara (2016) provide useful examples of including both symmetric tests with NHST findings and asymmetric tests with SPOT findings in the same studies.

Such parallel presentations of alternative research paradigms bring to mind the ancient (1997–2001) co-practices of submitting a paper for journal publication consideration by mailing paper copies with a cover letter to an editor via postal services while at the same time sending the paper as an email attachment to the same editor. Overlapping time periods occur in the use of alternative old and new technologies, theories, and data analysis. The overlap in the use of shallow symmetric tests of variable-directional relationships and deeper asymmetric tests of case-based complex antecedent conditions indicating precise outcomes is occurring in the present decade. Given the explosion of research offering asymmetric theories and tests, Hubbard's (2016) identification of NHST as corrupt research, and the contributions by Ragin (2000, 2008), Gigerenzer (2001), Zadeh (1965, 1996, 2010), and others (Hsiao, Jaw, Huan & Woodside, 2015; Nagy et al. 2017; Wu et al. 2014), the overlap is likely to end sometime during 2025–2030. Consequently, LaPlaca's calling for a true new paradigm shift in B-to-B research that includes useful, accurate predictions of complex outcomes will be realized. We are moving toward having our cake and eating it too – deep description and explanation as well as accurate prediction via asymmetric models.

REFERENCES

Ang, H. B. (Andy), & Woodside, A. G. (2017). Is Bart Simpson offering sage advice? A case-based general model of managers' core self-evaluations and job satisfaction. *Journal of Business Research*, *74*, 11–37.

Anterasian, C., Graham, J. L., & Money, R. B. (1996). Are U.S. managers superstitious about market share? *Sloan Management Review*, *37*, 67–77.

Armstrong, J. S. (2009). Peer review for journals: Evidence on quality control, fairness, and innovation. *Science and Engineering Ethics*, *3*, 63–84.

Armstrong, J. S. (2012). Illusions in regression analysis. *International Journal of Forecasting*, *28*, 689–694.

Armstrong, J. S., & Collopy, F. (1996). Competitor orientation: Effects of objectives and information on managerial decisions and profitability. *Journal of Marketing Research*, *33*, 188–199.

Armstrong, J. S., & Green, K. C. (2007). Competitor-oriented objectives: The myth of market share. *International Journal of Business*, *12*, 115–134.

Armstrong, J. S., & Hubbard, R. (1991). Does the need for agreement among reviewers inhibit the publication of controversial findings? *Behavioral and Brain Sciences*, *14*, 136–137.

Banerjee, A. V., & Duflo, E. (2008). *The experimental approach to development economics*. Working Paper No. 14467. National Bureau of Economic Research, Cambridge, MA. Retrieved from http://www.nber.org/papers/w14467.pdf. Accessed on May 30, 2017.

Bass, F. M., Tigert, D. J., & Lonsdale, R. T. (1968). Market segmentation: Group versus individual behavior. *Journal of Marketing Research*, *5*, 264–270.

Brenes, E. R., Ciravegna, L., & Woodside, A. G. (2017). Constructing useful models of firms' heterogeneities in implemented strategies and performance outcomes. *Industrial Marketing Management*, *62*, 17–35.

Cohen, J. (1994). The Earth is round ($p < .05$). *American Psychologists*, *49*, 1304–1312.

Cyert, R. M., Simon, H. A., & Trow, D. B. (1956). Observation of a business decision. *Journal of Business*, *29*, 237–238.

Duflo, E., Kremer, M., & Robinson, J. (2008). How high are rates of return to fertilizer? Evidence from field experiments in Kenya. *American Economic Review*, *98*, 482–488.

Feurer, S., Baumbach, E., & Woodside, A. G. (2016). Applying configurational theory to build a typology of ethnocentric consumers. *International Marketing Review*, *33*, 351–375.

Fiss, P. C. (2007). A set-theoretic approach to organizational configurations. *Academy of Management Review*, *32*, 1180–1198.

Fiss, P. C. (2011). Building better casual theories: A fuzzy set approach to typologies in organizational research. *Academy of Management Journal*, *54*, 393–420.

Frösén, J., Luoma, J., Jaakkola, M., Tikkanen, H., & Aspara, J. (2016). What counts versus what can be counted: The complex interplay of market orientation and marketing performance measurement. *Journal of Marketing*, *80*, 60–78.

Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, *98*, 254–267.

Gigerenzer, G. (2001). The adaptive toolbox. In G. Gigerenzer & S. Reinhard (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 37–50). Cambridge, MA: MIT Press.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristics: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*, 107–143.

Gladwin, C. H. (1989a). Indigenous knowledge systems, the cognitive revolution, and agricultural decision making. *Agriculture and Human Values*, *6*, 32–41.

Gladwin, C. H. (1989b). *Ethnographic decision tree modeling*. Beverly Hills, CA: Sage.

Gough, A. E., Gladwin, C. H., & Hildebrand, P. E. (2002). Vouchers versus grants of Inputs: Evidence from Malawi's starter pack program. *African Studies Quarterly*, *6*, 201–222. Retrieved from http://www.africa.ufl.edu/asq/v6/v6i1-2a8.pdf. Accessed on May 30, 2017.

Gummesson, E. (2006). Qualitative research in management: Addressing complexity, context and persona. *Management Decision*, *44*, 167–179.

Hadjikhani, A., & LaPlaca, P. (2013). Development of B2B marketing theory. *Industrial Marketing Management*, *42*, 294–305.

Howard, J. A., & Morgenroth, W. M. (1968). Information processing model of executive decision. *Management Science*, *14*, 416–428.

Hsiao, J. P. H., Jaw, C., Huan, T. C., & Woodside, A. G. (2015). Applying complexity theory to solve hospitality contrarian case conundrums: Illuminating happy-low and unhappy-high performing frontline service employees. *International Journal of Contemporary Hospitality Management*, *27*, 608–647.

Hubbard, R. (2016). *Corrupt research: The case for reconceptualizing empirical management and social science*. Los Angeles, CA: Sage.

Huff, A. S., Huff, J. O., & Barr, P. (2001). *When firms change direction*. New York, NY: Oxford University Press.

Judge, T. A., Locke, E. A., & Durham, C. C. (1997). The dispositional causes of job satisfaction: A core evaluations approach. *Research in Organizational Behavior*, *19*, 151–188.

Kotler, P. (1967). *Marketing management: Analysis, planning, and control*. Englewood Cliffs, NJ: Prentice-Hall.

LaPlaca, P., & da Silva, R. V. (2016). B2B: A paradigm shift from economic exchange to behavioral theory: A quest for better explanations and predictions. *Psychology & Marketing*, *33*, 232–249.

LaPlaca, P. J. (1997). Contributions to marketing theory and practice from industrial marketing management. *Journal of Business Research*, *38*, 179–198.

McClelland, D. C. (1998). Identifying competencies with behavioral-event interviews. *Psychological Science*, *9*, 331–339.

McCloskey, D. (2002). *The secret sins of economics*. Chicago, IL: Prickly Paradigm Press.

Misangyi, V. F., Greckhamer, T., Furnari, S., Fiss, P. C., Crilly, D., & Aguilera, R. (2017). Embracing causal complexity: The emergence of a neo-configurational perspective. *Journal of Management*, *20*, 1–28.

Montgomery, D. F. (1975). New product distribution: An analysis of supermarket buyer decisions. *Journal of Marketing Research*, *12*, 255–264.

Morgenroth, W. M. (1964). A method for understanding price determinants. *Journal of Marketing Research*, *1*, 17–26.

Nagy, G., Megehee, C. M., Woodside, A. G., Laukkanen, T., Hirvonen, S., & Reijonen, H. (2017). Achieving requisite variety in modeling firms' strategy heterogeneities: Explaining paradoxical firm-market performances. *Industrial Marketing Management*, *65*, 100–128.

Nelder, J. A. (1986). Statistics, science and technology (with discussion). *Journal of the Royal Statistical Society, Series A*, *149*, 109–121.

Ordanini, A., Parasuraman, A., & Rubera, G. (2014). When the recipe is more important than the ingredients: A qualitative comparative analysis (QCA) of service innovation configurations. *Journal of Service Research*, *17*, 134–149.

Phillips, L. A. (1968). *An exploratory study of factors which determine the initial distribution of selected categories of products on supermarket shelves*. Unpublished Master's Thesis, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA.

Powell, T. C., Lovallo, D., & Fox, C. (2011). Behavioral strategy. *Strategic Management Journal*, *32*, 1369–1386.

Ragin, C. C. (2000). *Fuzzy-set social science*. Chicago, IL: University of Chicago Press.

Ragin, C. C. (2006). The limitations of net-effects thinking. In B. Rihoux & H. Grimm (Eds.), *Innovative comparative methods for policy analysis*. Boston, MA: Springer.

Ragin, C. C. (2008). *Redesigning social inquiry: Fuzzy sets and beyond*. Chicago, IL: University of Chicago Press.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory resting. *Psychological Review*, *107*, 358–367.

Roig-Tierno, N., Gonzalez-Cruz, T. F., & Llopis-Martinez, J. (2017). An overview of qualitative comparative analysis: A bibliometric analysis. *Journal of Innovation & Knowledge*, *2*, 15–23.

Rumelt, R. (2011). *Good strategy, bad strategy*. London: Profile Books.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods*, *1*, 115–129.

Sterman, J. D., & Wittenberg, J. (1999). Path dependence, competition, and succession in the dynamics of scientific revolution. *Organization Science*, *10*, 322–341.

Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology*, *37*, 1–2.

Tucker, C., & Zhang, J. (2010). Growing two-sided networks by advertising the user base: A field experiment. *Marketing Science*, *29*, 805–814.

Tukey, J. W. (1991). The philosophy of multiple comparisons. *Statistical Science*, *6*, 100–116.

Urry, J. (2005). The complexity turn. *Theory, Culture and Society*, *22*, 1–14.

Wilson, R. D. (2010). Using clickstream data to enhance business-to-business website performance. *Journal of Business & Industrial Marketing*, *25*, 177–187.

Woodside, A. G. (2013). Moving beyond multiple regression analysis to algorithms: Calling for a paradigm shift from symmetric to asymmetric thinking in data analysis and crafting theory. *Journal of Business Research*, *66*, 463–472.

Woodside, A. G. (2014). Embrace•perform•model: Complexity theory, contrarian case analysis, and multiple realities. *Journal of Business Research*, *67*, 2495–2503.

Woodside, A. G. (2016). *Bad to good: Achieving high quality and impact in your research*. Bingley: Emerald.

Woodside, A. G. (2017). *The complexity turn: Cultural, management and marketing applications*. Berlin: Springer.

Woodside, A. G., Schpektor, A., & Xia, X. (2013). Triple sense-making of findings from marketing experiments using the dominant variable based-logic, case-based logic, and isomorphic modeling. *International Journal of Business and Economics*, *12*, 131–153.

Wu, P. L., Yeh, S. S., Huan, T. C., & Woodside, A. G. (2014). Applying complexity theory to deepen service dominant logic: Configural analysis of customer experience-and-outcome assessments of professional services for personal transformations. *Journal of Business Research*, *67*, 1647–1670.

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, *8*, 338–353.

Zadeh, L. (1996). Fuzzy logic: Computing with words. *IEEE Transactions on Fuzzy Systems*, *4*, 103–111.

Zadeh, L. (2010). Lofti Zadeh on CWW. In J. M. Mendel, L. A. Zadeh, E. Trillas, R. Yager, J. Lawry, H. Hagras, & S. Guadarrama (Eds.), *What computing with words means to me*. *IEEE Computational Intelligence Magazine*, February, 20–26. Retrieved from https://people.eecs.berkeley.edu/~zadeh/papers/What%20Computing%20with%20Words%20Means%20to%20Me-CIM%202010.pdf. Accessed October 8, 2016.

Zellner, A. (2001). Keep it sophisticatedly simple. In H. Keuzenkamp & M. McAleer (Eds.), *Simplicity, inference, and modelling: Keeping it sophisticatedly simple*. Cambridge: Cambridge University Press.