How to move away from using symmetric tests, net effects, and $p < 0.05$

Overcoming barriers to good science practices

Arch George Woodside

Coastal Carolina University, Conway, South Carolina, USA

Abstract

Purpose – The purpose of this paper is to describe how and why to shift away from bad science practices now dominant in research to good science practices.

Design/methodology/approach – The essay includes details in theory construction and the use of symmetric tests to illustrate bad science practices. In contrast, the essay includes asymmetric case-based asymmetric theory construction and testing to illustrate good science practices.

Findings – Researchers in marketing science should not report null hypothesis significance tests. They should report somewhat precise outcome tests, avoid using multiple regression analysis (MRA) and do use Boolean-algebra-based algorithms to predict cases of interest.

Research limitations/implications – Given the widespread dominance of bad science practices (e.g. MRA and structural equation modeling), the inclusion of both bad and good science practices may be necessary during the transition years of 2015–2025 (e.g. Ordanini et al., 2014).

Practical implications – Good science practices fit reality much closer than bad science practices. Asymmetric modeling includes recognizing the separate models are necessary for positive vs negative outcomes because the antecedents of each often differ.

Originality/value – This essay presents details of why and how researchers need to embrace a new research paradigm that is helpful for ending bad science practices that are now dominant in research in marketing.

Keywords Symmetric, Asymmetric, Case-based, Good science practice

Paper type Conceptual paper

Introduction: gaining perspective on pervasive bad science practices

Hubbard (2015, pp. 194-198) reviews 41 “overt criticisms of the worth of null hypothesis significance testing (NHST).” These criticisms include Hunter’s (1997) call for banning the practice: “Needed: a ban on the significance test.” The editor-in-chief of Basic and Applied Social Psychology has done just that:

The Basic and Applied Social Psychology (BASP) 2014 Editorial emphasized that the null hypothesis significance testing procedure (NHSTP) is invalid, and thus authors would be not required to perform it (Trafimow, 2014). However, to allow authors a grace period, the Editorial stopped short of actually banning the NHSTP. The purpose of the present Editorial is to announce that the grace period is over. From now on, BASP is banning the NHSTP (Trafimow and Marks, 2015).

While Hubbard (2015) and Trafimow and Marks (2015) provide details and convincing evidence that the use of null hypothesis significance testing (NHST) is bad science practice, Tukey (1991, p. 100) provides the obvious (but still shocking) conclusion: “All we know about the world teaches us that the effects of A and B are always different – in some decimal place – for any A and B. Thus, asking, ‘Are the effects different?’ is foolish.”

Following this introduction, the second section of this essay documents details in the literature supporting the paradigm shift from variable-based to case-based model construction and testing. The third section describes the “the forces of inertia” (Huff et al., 2001) and barriers causing the slow adoption process of modeling precise outcomes and discarding NHST. The fourth section includes visuals of the causal configurations of
antecedents supporting the current pervasive use of bad vs good science practice in empirical sub-disciplines of business and the behavioral sciences (e.g. finance, management, marketing, psychology and tourism/hospitality research). The fifth section describes core principles of good science practices appearing in a few articles, but still infrequently, in journals in these literatures. To stimulate the shift from using corrupt research practices to good science practices, the fifth section briefly compares the use of bad and good science practices with the same data set in two articles. After recognizing the opportunity, for researchers seeking to leave behind the shallow waters of NHST and other corrupt research practices and dive deeply into achieving requisite variety via asymmetric theory construction and modeling, the sixth section concludes with the suggestion to read a few selections from the literature on good science practices.

Weick (2007, p. 16) provides a useful introduction on the need to achieve requisite variety that is relevant for recognizing the need to embrace complexity theory and case-based modeling: “The importance of a head full of theories is that this increases requisite variety. By that I mean that it takes a complicated sensing device to register a complicated set of events. And a large number of theories can be a complex sensing device if believing is seeing. Haberstroh (1965, p. 1176) describes the law of requisite variety this way: ‘If the environment can disturb a system in a wide variety of ways, then effective control requires a regulator that can sense these disturbances and intervene with a commensurately large repertory of responses.’ Thus, it takes richness to grasp richness.” The current dominant logic of reporting one-to-five regression models in symmetric tests is too simplistic of an approach to the rich tapestry inside most data files. While parsimony is a worthy objective in data analysis, nearly all symmetric tests are overly simplistic.

**Documenting the reasons for embracing the paradigm shift from variable-based to case-based modeling**

This section describes the configuration of conditions supporting the rejection of good science practices that include statistical sameness testing (somewhat precise outcome testing (SPOT)) and additional good science practices (e.g. model construction that recognizes and attempts to explain/predict anomalies.) Figure 1 illustrates the configuration of rejection conditions and the conditions of good science practices rarely appearing in practice.

Hubbard (2015, p. 9) offers substantial convincing evidence supporting his conclusion that NHST is corrupt research: “In a nutshell, this book demonstrates that the significant difference paradigm is philosophically suspect, methodologically impaired, and statistically broken.”

---

**Figure 1.** Forces of inertia and barriers preventing shifting from bad (“corrupt research,” NHST) to good science (SPOT)

- Forces of inertia: high comfort with status quo of MRA/SEM model construction
- Fear of rejection by reviewers and editors if NHST theory and testing not used
- Implicit/explicit knowledge editors and reviewers prefer to reject controversial findings
- Unawareness and lack of training in modeling and testing by algorithms

**Adoption of good science practices**

- SPOT: Somewhat precise outcome testing
- Use of XY plots of data outputs
- Predictive validation of models via additional samples
- Anomaly theory construction (4 corners)
- Ex ante theory construction of specific complex antecedent conditions
- Achieving requisite variety: Embracing complexity theory tenets
Demonstrating that “[…] a difference between two means is not precisely zero, or that a correlation between to variables is not precisely zero, are trivial findings” (Cohen, 1994, p. 1000). Hubbard (2015, pp. 192-193) points out that empirical management and behavior science (EMBS) researchers can do better: “In principle, there is no reason why theories in the management and social science 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.” Though NHST analytics (e.g. multiple regression analysis (MRA) and structural equation modeling (SEM)) dominates across recent decades, studies that include point prediction analytics are available in the relevant EMBS literatures (e.g. Gigerenzer, 1991; Gigerenzer and Brighton, 2009; McClelland, 1998; Montgomery, 1975; Morgenroth, 1964). A core point here is that precise predictions are bounded inside contexts – the idea that the natural sciences, physics included, deal with phenomena that are not context dependent is a myth (Holzman, 1986, p. 348; Hubbard, 2015, p. 82). Context dependence of precise predictions follows also from Simon’s (1990, p. 1) scissors metaphor: “Human rational behavior is shaped by a scissors whose blades are the structure of task environments and the computational capabilities of the actor.” For useful model construction, given that the shaping of precise outcomes by two forces – task environment (context) and actor capabilities and backgrounds – is the first axiom, the second axiom is that accomplishing accurate predictions of precise outcomes requires identifying configurations that include combinations of the context and actor features. Granting that including a combination of several context-actor features restricts the range of generalization does not negate the conclusions that researchers can estimate precise outcomes accurately and that testing the accuracy of alternative configurations and with additional features and reductions in the number of features are steps toward generalizing models of precise outcomes.

Somewhat hidden in this discussion of shifting from NHST directional predictions to predicting precise outcomes is the inherent shifting from early discussions in research articles from case-based discussions to writing variable-based hypothesis and examining the existence of relationships via analytics using continuous variable data. Followed at the end of most articles using the dominant logic of variable-based, NHST, to a shift back to presenting implications at the case-based level. Fiss (2007, p. 1181) tellingly describes this three-step awkward shifting: “But while theoretical discussions of configurational theory thus stress nonlinearity, synergistic effects, and equifinality, empirical research has so far largely drawn on econometric methods that by their very nature tend to imply linearity, additive effects, and unifinality. This mismatch has caused a number of problems. For example, the classic [still dominant] linear regression model treats variables as competing in explaining variation in outcomes rather than showing how variables combine to create outcomes. By focusing on the relative importance of rival variables, a correlational approach has difficulty treating cases as configurations and examining combinations of variables. This becomes particularly evident in the fact that regression analysis focuses on the unique contribution of a variable while holding constant the values of all other variables in the equation.” Fiss (2007, p. 2007) concludes: “Set-theoretic [case-based] approaches are particularly adept at identifying localized effects. Rather than estimating the relative importance of different strategies across all cases, set-theoretic methods allow us to better examine which strategies make sense for which kinds of firm. By contextualizing effects, it becomes easier to go beyond global and typically vague statements about effects, and the identification of different paths rather than a single path offers more opportunities for policy intervention (Ragin and Fiss, 2007).”

The currently dominant paradigm stance asks if an XY relationship is significantly different from zero. The latter and new paradigm asks, what configurations (i.e. screens) of conditions lead to a given outcome, for example, forecasting stock price growth by 10 percent plus for firms in industry X in the top quintile across each of five financial/
marketing metrics (bottom quintile on price/equity ratio, top quintile on sales recent year sales growth, top quintile on customer satisfaction and so on) is case-based model approach – using a configurational screen rather than a variable-based regression analysis. Case-based predictive modeling is applying asymmetric tests using Boolean algebra; variable-based predictive modeling is applying symmetric tests using matrix algebra. While heretofore unrecognized as a substantial change, the shifts from using symmetric, variable-based tests via regression analysis to asymmetric, case-based tests using algorithms (i.e. screens) by Montgomery (1975) and McClelland (1998) are earthquakes. Both Montgomery (1975) and McClelland (1998) described their use of both symmetric and asymmetric tests; the resulting meager amount of useful information from symmetric tests was their rationale for reporting findings for case-based asymmetric tests. Thus, without their stating the fact, Montgomery (1975) and McClelland (1998) shifted from using corrupt research tools and bad science practices to using honest research tools and good science practices. Subsequently, years later, additional studies include both paradigms with examples of the same shift from the dominant variable-based to the new case-based prediction paradigm in different EMBS contexts (Ferguson et al., 2017; Frösén et al., 2016; Ordanini et al., 2014). Using less critical rhetoric than appearing in Hubbard (2015) and Woodside (2016a, b), these three additional studies have direct comparisons of findings from symmetric and asymmetric testing. Additional studies (Fiss, 2007, 2011; Hsiao et al., 2015; Woodside, 2013; Wu et al., 2014) explicitly call attention to the benefits from embracing the case-based modeling paradigm and the greater usefulness of asymmetric (precise outcome) tests.

Why bad science practices pervade empirical management and social science

Hubbard (2015) describes the forces of inertia and barriers to shifting to the superior research paradigm of “statistical sameness testing” (what this present essay refers to as “SPOT”) – the resulting predictions in case-based modeling are somewhat precise because an accuracy hit ratio of 100 percent is rarely obtainable. “What is the likelihood of the statistical sameness paradigm supplanting, or at least, paralleling, that of significant difference? This is a tall challenge, one made worse by two formidable and interrelated barriers. The first is that members of the significant difference paradigm cherish their conception of science and tend to look at alternative approaches to knowledge procurement with an air of suspicion, or even dismissiveness […] [They] wish to keep doing things the same way they have been done for decades. As a consequence, the sheer weight of academic inertia, fortified by researcher unawareness of how science makes headway, acts as a powerful antidote against the need for change of any kind. The second barrier is that all too often academicians are preoccupied with enriching their careers as scholars and are unconcerned with real-world knowledge development” (Hubbard, 2015, p. 228).

Related to the second barrier are pedagogical assessments that indicate acceptance only of the currently dominant paradigm. The following statements actually expressed by full professors reflect such reasoning. “SEM is standard practice. Everybody uses SEM, so I must do so.” “I can’t change [instruction content in the marketing research course] just because it’s wrong.” The second quote was supported by the following statement: “I don’t have the time in the course schedule to teach SPOT.” Fear of rejection is viewable as a separate causal condition. Protests against SPOT from the significant difference school do not center on that NHST is wrong but on concerns that graduate students would not be able to publish their work unless the work used statistical significance testing (Hubbard, 2015; Schmidt, 1996).

For early career scholars, the low right condition in the Venn diagram in Figure 1 may be in for a surprise. However, extensive evidence on peer review shows that papers with findings that contradict important viewpoints are nearly always rejected by reviewers (Armstrong, 1997). For example, a survey by Armstrong and Hubbard (1991) found
Editors of 16 psychology journals reported that reviewers dealt harshly with papers that contained controversial findings. Armstrong [...] found that none of what he considers his twenty most important papers received full acceptance by reviewers (Armstrong and Green, 2007). Adopting a stance somewhat hiding your true purpose via: “I come here to bury Caesar, not to praise him” – followed by accurate evidence contrary to this verbal statement is more likely to be successful than trying a full-frontal attack on a dominant paradigm. Reading Ordanini et al. (2014) might spring-to-mind as a successful execution this strategy. Gigerenzer and Brighton (2009) illustrate a full-frontal attack on regression analysis with a rich and deep set of evidence of good science practices (see also Armstrong, 2012).

The start-ups of multiple catalytic actions on several fronts are often necessary to gain widespread adoption of a superior new technology (Woodside, 1996). Part of the solution for breaking through the barriers preventing good science practices surely is the approach by Montgomery (1975), McClelland (1998), Ordanini et al. (2014) and Ferguson et al. (2017) of including theory, data analysis and findings using both NHST and SPOT. The outright banning of NHST reporting (Trafimow and Marks, 2015) represents a draconian solution. Describing and illustrating the findings and benefits of the new paradigm appearing in articles in elite journals (Frösén et al., 2016; Prado and Woodside, 2015) represent a third catalyst. Joining and becoming active in the world’s leading organization on case-based modeling and estimating the accuracies of precise outcome predictions – COMPASSS.ORG – are a fourth catalyst. Participating in workshops to learn how to construct and test theory using algorithms (fuzzy-set qualitative comparative analysis) sponsored, for example, by COMPASSS.ORG, Global Innovation and Knowledge Academy and the Global Alliance of Marketing and Management Associations is a fifth catalyst.

Visualizing bad vs good science practices
This section presents and describes visual outputs of bad and good science practices. Gigerenzer’s (1991, p. 19) wisdom has great importance here: “Scientists’ tools are not neutral.” The current dominant logic in EMBS is bad science practice, that is, constructing theories from a foundation of regression analysis (MRA/SEM) – symmetric tests of relationships between variables showing differences are not equal to zero. The new logic in EMBS is good science, which is, constructing theories form a foundation of algorithms – asymmetric tests of complex antecedent screens indicating consistent occurrence of a specific outcome. The examples of bad vs good practices in Figure 2 make use of the data from the same study.

Figure 2 presents a variable-based theory of antecedents and consequences for problem gambling (PG) from a study by Prentice and Woodside (2013). Part A in Figure 2 presents the theory. Part B in Figure 2 presents the findings. Symmetric testing using MRA is the basis for both the theory construction and data analysis. The core hypotheses are directional predictions (e.g. as age increases, PG increases). Part A shows individual positive and negative associations for nine antecedents and PG: five casino customer socio-economic status characteristics and four customer behaviors. The hypotheses are only directional and, as such, illustrate shallow theory and testing. As McCloskey (2002) and Hubbard (2015) emphasize, without a focus on quantities (i.e. precise outcomes) such theorizing are “worthless as science” (McCloskey, 2002, p. 55). Using directional hypotheses illustrates shallow research practice and disinformation. Such research lacks the “requisite variety” to describe, explain and predict anomalies in relationships – and anomalies almost always occur in relationships.

For example, while age may have a positive association with casino PG, a number of young people are likely to be casino problem gamblers. (Note the shift from “problem gambling” to “problem gambler.”) Discretizing cases (individual respondents’ data) by age and severity of PG using quintiles results in 25 cells and with reasonably large sample sizes (n ≥ 100); all 25 cells include a few to many cases even when the symmetric test indicates a highly significant positive
Notes: (a) Part A: theory; (b) Part B: variable findings. *p<0.05
Source: Prentice and Woodside (2013, p. 1110)
relationship. Just reporting that the majority of cases are found in the main diagonal (young and not PG vs old and PG) is shallow bad science. Young-and-PG cases and old-and-not-PG cases are not unexplainable blips but seeming anomalies worthy of theory construction and testing. “Four-corner theory construction” is the recognition that most XY associations for a simple X and a simple Y support the complexity theory tenet (Woodside, 2017) that all four associations occur: low X and high Y, high X and high Y, high X and low Y and low X and low Y— even when the effect size of the directional hypothesis between X and Y is large.

An anomaly is a fact or case that does not fit received wisdom. “To 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. Shifting from variable-based to cases-based theory construction and data analysis are steps necessary to take to fully examine anomalies. Discretizing using quintiles (McClelland, 1998) or fuzzy-set scores (Ragin, 2008) are adequate procedures for shifting successfully. Because cases tend to clump around the median and the cases in the bottom and top quintiles are actors of particular interest, dichotomizing cases into low and high scores is a bad practice in transitioning from variable-based to case-based modeling (cf. Fitzsimons, 2008). Iacobucci et al. (2015a, b) offer an alternative and incorrect perspective that supports the bad science practice of dichotomizing data into cases with low vs high scores for a given variable. Except for naturally occurring dichotomous variables (e.g. gender), researchers should avoid dichotomizing continuous variables as Rucker et al. (2015) recommend. However, Rucker et al. (2015) are mistaken and offer bad advice in recommending that preserving the continuous nature of the variable and analyzing the data via linear regression and in recommending that regression remain the normative procedure in research involving continuous variables.

The findings in Part B of Figure 2 include simple standardized regression weights testing the null hypotheses of zero associations. Note that the findings support the pattern of positive and negative hypotheses for the antecedent conditions to PG but not the positive hypotheses for the consequences of PG. Rather than positive relationships, the three consequences from PG to evaluating casino services are negative (as well as for the additional services shown in Part A and included in Part B). This pattern of findings leads Prentice and Woodside (2013) to entitle their article: “Problem gamblers’ harsh gaze on casino services.”

However, Prentice and Woodside (2013) overgeneralize and their theory and data analysis remain in the shallows. Not all problem gamblers gave a negative assessment of casino services as their study suggests. Shifting from a variable-based to a case-based theory and data analysis indicates that 20 percent of customers high in PG provided positive assessments of casino services (Woodside et al., 2015). By discretizing using quintiles and constructing and testing a four-corners’ theory, the research leaves the shallows and adopts a requisite variety perspective—a deep dive in understanding, describing and predicting problem gamblers as well as non-problem gamblers and the complex antecedent conditions indicating each of the four corners.

Figure 3 shows the cases in the corners for PG and casino service assessments. While the distribution of shares of cases indicates a negative PG-assessment association in Figure 3, cases appear in all 25 cells and this finding supports asking a series of “who” rather than “if” questions. Who are the problem gamblers giving positive vs negative casino assessments? Who are the non-problem gamblers giving positive vs negative casino assessments? The findings in Figure 4 permit answering these questions. Figure 4 provides asymmetric models—predicting precise outcomes in one direction such as customers providing highly positive assessments of casino services (Part A in Figure 4) and customers providing highly negative assessments of casino services (Part B in Figure 4).
Contrarian type 2 cases: 39% of customers with zero to very low problem-gambling scores gave low scores on overall service quality.

30% of customers with zero to very low problem-gambling scores gave high scores on overall service quality.

48% of customers with high to very high problem-gambling scores gave low scores on overall service quality.

Contrarian type 1 cases: 20% of customers with high to very high problem-gambling scores gave high scores on overall service quality.

Notes: For the distribution of cases, the symmetric main effect is negative; $\varphi = 0.288$, $p < 0.081$. ANOVA findings indicate significant differences in overall service quality by problem gambling segments that support a significant symmetric negative main effect, means (standard errors) for the five PG segments from low to high: 9.82 (0.10); 9.31 (0.21); 9.67 (0.19); 9.66 (0.24); 9.05 (0.26); $F = 2.68$, df = 4/406, $p < 0.032$. The findings include contrarian type 1 cases: cases with high scores on the outcome condition that counters the negative symmetric main effect; the findings include contrarian type 2 cases: cases with low scores on the outcome condition that counters the negative symmetric main effect.

Source: Woodside et al. (2015, p. 68)

Using symmetric tests

Part A: Four complex antecedent conditions indicate highly positive overall casino assessments

<table>
<thead>
<tr>
<th>Model</th>
<th>PGSI</th>
<th>Age</th>
<th>Gender</th>
<th>Education</th>
<th>Income</th>
<th>Occ Status</th>
<th>Title Status</th>
<th>Coverage</th>
<th>Raw</th>
<th>Unique</th>
<th>Consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>0.16</td>
<td>0.10</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>0.14</td>
<td>0.06</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>0.05</td>
<td>0.94</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>0.08</td>
<td>0.02</td>
<td>0.84</td>
<td></td>
</tr>
</tbody>
</table>

Solution coverage = 0.30; solution consistency = 0.82

Part B: Eight complex antecedent conditions indicate highly negative overall casino assessments

<table>
<thead>
<tr>
<th>Model</th>
<th>PGSI</th>
<th>Age</th>
<th>Gender</th>
<th>Education</th>
<th>Income</th>
<th>Occ Status</th>
<th>Title Status</th>
<th>Coverage</th>
<th>Raw</th>
<th>Unique</th>
<th>Consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>0.27</td>
<td>0.05</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>0.13</td>
<td>0.03</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>0.09</td>
<td>0.02</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>0.15</td>
<td>0.03</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>0.15</td>
<td>0.05</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>0.04</td>
<td>0.01</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>0.17</td>
<td>0.00</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>0.13</td>
<td>0.00</td>
<td>0.79</td>
<td></td>
</tr>
</tbody>
</table>

Solution coverage = 0.50; solution consistency = 0.83

Notes: PGSI, problem gambling severity index; “●”, high score; “ο”, low score; “(blank)” not relevant to the model

Source: Woodside et al. (2015, p. 75)
The models in Figure 4 are expressible by “computing with words” (Zadeh, 1996). For example, model 1 in Part A in Figure 4 states that old female casino guests, low in education and income, high in occupational status, who are not problem gamblers, provide high positive assessments of casino services. However, the consistency index (0.76) for this prediction indicates a number of exceptions occur. The consistency index in asymmetric analysis is analogous to a correlation ($r$) and the coverage index is analogous to the “coefficient of determination” ($r^2$). Ragin (2008) provides details for computing consistency and coverage indexes. Woodside (2017) recommends that models achieving very high consistencies (above 0.85) are particularly useful for indicating sound theory and information for practice.

Principles of good science practice that are usually missing in most articles

Six good science practices that are usually missing in EMBS articles in elite and lower ranked journals appear inside the right-side circle in Figure 1. Woodside (2016a, b) provides additional discussion of these and additional good practices usually missing in research reports in EMBS journals.

SPOT is the first good science practice in the right-side of Figure 1. This present essay and Hubbard (2015) in particular explain why researchers should shift to using SPOT and leave the corrupt research practices of NHST behind. Reporting XY plots is very useful practice; Anscombe (1973) demonstrates that very different XY plots can occur for different data sets having the same mean, standard deviation and correlation. Even though Anscombe has one thousand plus citations via Google.com/scholar, the practice of including XY plots occurs rarely.

Second, studies in articles in leading journals frequently report statistically significant fit validities and no analysis for predictive validation of models using separate samples. Armstrong (2012) demonstrates accomplishing significant fit validity using a table of random numbers. Reporting only fit validity is bad science practice as Armstrong (2012), Gigerenzer and Brighton (2009), McClelland (1998), Morgenroth (1964) and Roberts and Pashler (2000) all stress. Fit validity indicates that symmetric tests outperform asymmetric tests in accuracy; predictive validation supports the opposite conclusion (Gigerenzer and Brighton, 2009). The “critical issue is whether or not a model is useful in practice that is, does the model have high predictive validity when testing on additional samples not used in constructing the theories” (McClelland, 1998, p. 335). This perspective is known widely but rarely practiced in research reports in scholarly journals.

Third, theory construction for explaining, describing and predicting anomalies rarely appears in elite and lower ranked journals. Given that anomalies are recognizable in nearly all data files, this observation may be surprising. However, the pervasive practice of bad science focusing on NHST via symmetric tests prevents scholars from diving deep into identifying and modeling anomalous cases to statistically significant directional relationships. Embracing complexity theory (Woodside, 2017) and case-based theory construction and testing will cause a substantial increase in the study of anomalies.

The fourth good practice is to use algorithms to test for complex outcome configurations. Similar to Armstrong’s (2012) discussion on the bad practice of using stepwise regression analysis (like playing tennis without a net; something is bound to be significant if you include 8–25 terms in regression model), algorithm software permits the researcher to test for whatever complex configurations will indicate a precise outcome consistently. Researchers should include ex ante model construction of algorithms rather than just seeing what models the software produces. Ferguson et al. (2017) elaborate on this tenet of good science practice.

Embracing tenets of complexity theory (Woodside, 2014, 2017) – the fifth good practice – nurtures the achievement of requisite variety, shifting to asymmetric from symmetric tests, moving away from NHST to SPOT and leaving the shallows in EMBS. Complexity theory includes the following tenets. 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 of cases (e.g. 7 of 8, 14 of 15, 25 of 27), but a few false positives occur – thus, the expression “SPOT.” 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 identifies 10–20 percent 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 research.

Concluding remarks: readings and constructing theory away from the shallows
In a study of the impact of articles appearing during 2004–2008 in the Journal of Consumer Research, Pham (2013, p. 412) reports: “Very few articles – less than 10% – get very well cited, and the vast majority – roughly 70% – hardly ever get cited. In other words, the vast majority of the research that gets published, even in our top journals – perhaps 70% of it – hardly has any measurable scholarly impact in terms of citations.” Most of the journal articles in elite and lower ranked journals represent bad science via corrupt research practice. This conclusion follows from reading Gigerenzer and Brighton (2009), Hubbard (2015), Ragin (2008) and Woodside (2016c, 2017).

Read Hubbard (2015) to learn how bad science practice dominates today via symmetric NHST. Read Ragin (2008) to learn how to “redesign social inquiry” based on case-based theory construction and testing. Read Woodside (2017) for additional explanation of why embracing complexity theory is necessary for achieving requisite variety in theory construction and data analysis. During the current transition years from bad-to-good science practices, along with continuing to use bad science practices to break through the resistance barriers and forces of inertia supporting the dominant symmetric logic, report good science research practices as well in your research (e.g. Montgomery, 1975; Ferguson et al., 2017; Frösén et al., 2016; Ordanini et al., 2014).

References


Further reading


Corresponding author

Arch George Woodside can be contacted at: arch.woodside@bc.edu

For instructions on how to order reprints of this article, please visit our website:
www.emeraldtogrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com