Persuasion Principles Index: Ready for Pretesting Advertisements

Kesten C. Green

University of South Australia Business School and Ehrenberg-Bass Institute,

University of South Australia, Adelaide, Australia

Phone: +61 8 8302 9097; kesten.green@unisa.edu.au

J. Scott Armstrong

The Wharton School, University of Pennsylvania, Philadelphia, PA, U.S.A. and

Ehrenberg-Bass Institute, University of South Australia, Adelaide, Australia

Phone: +1 610-622-6480; armstrong@wharton.upenn.edu

Rui Du

University of Hawaii, Manoa, HI, U.S.A.
Phone: +1 215-275-6550; ruidu@hawaii.edu

Andreas Graefe

LMU Munich, Geschwister-Scholl-Platz 1, München, Germany

Phone: +49 89 21809466; a.graefe@lmu.de

Abstract

Purpose: To respond to issues posed in the four commentaries on Armstrong, Du, Green and Graefe (this issue) regarding the immediate usefulness of that paper's test of advertisements' compliance with persuasion principles, and regarding the need for further research.

Approach: Address commentators' concerns using logic, prior research findings, and further analyses of the data.

Findings: The superiority of the index method remains when a simple, theory-based, alternative weighting-scheme is used in the index model. Combinations of three unaided experts' forecasts were more accurate than the individual forecasts, but the gain was only one-third of the gain achieved by using the Persuasion Principles Index (PPI).

Research implications: Replications and extensions using behavioral data and alternative implementations of the index method would help to better assess the effects of judging conformity with principles as a means of predicting relative advertising effectiveness.

Practical implications: Advertisers can expect more accurate pretest results if they combine the predictions of three experts or, even better, if they use tests of compliance with persuasion principles, such as the PPI. The PPI software is copyrighted, but is available now and is free to use.

Originality/value: New analysis and findings provide further support for the claim that advertisers who use the PPI approach proposed by Armstrong, Du, Green and Graefe (this issue) to choose among alternative advertisements will be more profitable than those who do not.

Keywords: advertising effectiveness, combining forecasts, creativity, index method, profitability, variable weighting.

We are fortunate to have four leading researchers on communication, advertising, and marketing commenting on our tests of the Persuasion Principles Index (PPI) as a method for predicting the relative effectiveness of advertisements by assessing their compliance with scientific principles. They each raise distinct challenges, as one might expect when leading scholars with diverse perspectives comment on an approach that has not previously been proposed for advertising. We discuss the commentaries in the alphabetical order of the authors.

O'Keefe: Methodology Issues and Alternative Methods

O'Keefe (this issue) asks: (1) does our paper aim to test the predictive validity of the Armstrong (2010) principles and, if so, are there methodological issues; (2) was sufficient information about the research provided in the paper; and (3) can one use the PPI to create ads? These are important issues, and ones that other readers might also have considered.

(1) Does our paper aim to test the predictive validity of the principles? The first sentence in our paper stated, "To test whether a structured application of persuasion principles might help improve advertising decisions," while the opening sentence of the conclusion states, "This study provides a test of the predictive validity of persuasion principles." By the latter statement, we intended to communicate that we had tested the value of the body of relevant principles for selecting the more effective of a matched pair of advertisements.

O'Keefe points out that readers might interpret the statement in our paper's conclusion to mean that we had tested the predictive validity of each of the principles. He correctly assumes that we did not mean that. Consequently, the methodological concerns that he goes on to describe do not apply to our paper. As he states, "If the purpose is to provide proof-of-concept for the idea that a useful procedure is possible, this methodological concern is irrelevant."

O'Keefe's interest in how the predictive validity of individual principles should be tested is nevertheless important. We hope others will conduct such tests and contribute to the cumulative knowledge on persuasion. Developing and refining principles will never end, so researchers will likely discover evidence for additional principles. We also hope that researchers will obtain more evidence on existing principles: as Armstrong (2011) describes, only 22% of the 195 principles are based on "much experimental evidence," while 48% are based on "some experimental evidence," and 21% of principles

are supported by only one experimental study. Table 1 in Armstrong (2011) identifies ten principles in need of experimental research. The three most in need of further research are:

1. "Provoke customers only when it attracts attention to a selling point."
2. "Focus on benefits and features rather than choices when trying to change behavior."
3. "When the target market has an opposing viewpoint, consider using a story."

Thus, research to date on persuasion principles provides "not even the beginning of the end; but it is perhaps, the end of the beginning" (Churchill, 1942).

O'Keefe is arguably the world's leading researcher involved in aggregating evidence about the effects of alternative attempts at persuasion. In Armstrong (2010), there are 23 references to his many meta-analyses. O'Keefe's substantial contribution to the development of persuasion principles is due to his use of experimental studies as the primary source of evidence. We agree with O'Keefe (2015b) that valid principles are those that are based on experiments and replications designed to estimate effect sizes.

(2) Was sufficient information about the research provided in the paper?

We also agree with O'Keefe on the need for full disclosure to facilitate replication. However, we differ on the manner in which necessary information should be provided. O'Keefe suggests that all the information needed for a researcher to replicate a study should be contained in the article itself. We argue that disclosure should be conditional on the amount of material required for replication. In the case of our paper, the amount of information needed to replicate our research is too much to include in a journal article and would likely distract many readers who might otherwise find our paper useful. To help researchers, we have provided detailed information on the data and the PPI software in the Research Repository at advertisingprinciples.com (AdPrin.com). The Repository addresses all of the questions posed by O'Keefe. Over the three-plus years we have spent working on our paper, we have found that taking the one-hour self-administered training course and examining the PPI software are the best ways to understand the details of the procedure.

(3) Can one use the PPI to create ads?

O'Keefe raises the important issue of whether the persuasion principles can be used to *create* advertisements. Testing the effectiveness of using the principles to *create*

advertisements was not an objective of our paper. We agree that this would be a highly useful research project, no matter what the outcome. To test the usefulness of the principles for ad creation, a researcher could give expert advertisers a set of advertising design problems and ask them to create ads. Half of the experts could be randomly allocated to use the persuasion principles checklist on AdPrin.com, and the other half left free to follow their usual approach as the control group. Running the ads would provide evidence, in the form of predetermined measurable and useful criteria for success, on their relative effectiveness.

Sharp and Hartnett: Validity Threats and the Effects of Combining

Sharp and Hartnett (this issue)—hereafter S&H—suggest that our comparative tests might favor the PPI because we:

1. Used the *Which Ad Pulled Best* (*WAPB*) data;
2. Confined our tests to high-involvement, utilitarian products;
3. Used recall rather than sales as the objective criterion;
4. Tested accuracy against an unrepresentative benchmark;
5. Employed research subjects unrepresentative of the experts who would be retained in practice;
6. Failed to compare the accuracy of the PPI forecasts with the accuracy of combined forecasts from all of the other methods.

We address their concerns as follows.

(1) The effect of using WAPB data

S&H are correct that Armstrong's (2010) book, which introduced the persuasion principles that are the basis of the PPI, did use *WAPB* data. The principles were, however, developed without reference to the *WAPB* data or to studies that utilized that data. The *WAPB* data were used only to test the concurrent validity of the principles in the book. The principles were not changed in any substantive way in response to the concurrent validity test results. The process is described in Appendix B of Armstrong (2010) with further analyses in Armstrong and Patnaik (2011).

S&H point out that the evidence from Armstrong's (2010) concurrent validity testing using the *WAPB* data influenced the weighting of the principles in the PPI model. We did not expect that to be a threat to the validity of our tests, given that the weighting of variables in linear models tends not to have a substantial effect on forecast accuracy

(Graefe, 2015). Nevertheless, to address S&H's concerns, we designed a variation of the PPI that did *not* use evidence from *WAPB* in assigning weights to the variables.

We postulated that unequal weighting would be appropriate, as some of the principles would likely have stronger effects on persuasion than others. For example, there are only two persuasion principles under the heading of "1.1 Benefits," but 15 under the heading of "1.4 Price." Equal weighting of individual principles would imply that price is 7.5 times more important than benefits, which seems unlikely. Rather than assessing how much each principle should be weighted one at a time in ignorance of *WAPB* evidence, we employed the simplest weighting scheme that seemed logical in order to construct a theory-weighted PPI.

Specifically, we reasoned that the 85 headings (from "1.1 Benefits" to "10.4 Pace") are broadly equivalent in importance. That assumption implies that the weight assigned to an individual principle should be inversely proportional to the number of principles listed under the heading that are rated for the ad. For example, if both "5.10 Innuendos" principles were rated, they would each be weighted 0.5, whereas if all six "5.11 Customer Involvement" principles were rated, they would each be weighted 0.167. We refer to this alternative index model as the PPIx, with the "x" signifying that the *WAPB* evidence is excluded from the model. The construction of the PPIx is otherwise the same as that of the PPI.

The PPIx provided a forecast error reduction of 32% compared to forecasts from the copy-testing treatment. That error reduction is somewhat less than that from our original PPI, which was 37%, but it is still substantially greater than that from the alternative methods.

(2) The effect of testing only high-involvement utilitarian products

The experimental studies summarized in the Armstrong (2010) persuasion principles were conducted using persuasion problems in different disciplines and cultures. The principles are conditional and cover all conditions known to date. Consequently, we consider it reasonable to assume that the principles apply whenever persuasion is attempted.

We expected the principles to be more effective under certain conditions, specifically ads for high-involvement utilitarian products (e.g., automobiles) versus ads for low-involvement products (e.g., chewing gum). If our test had failed to find that PPI scores were substantially better than unaided judgment at picking the more effective ad

for high-involvement products, there would be little reason to test the PPI under other conditions. Following the same logic, we advise advertisers to initially apply the PPI where it will be most profitable—and where it has been tested.

The profitability of the PPI depends not only on the conditions, but also on the total advertising budget. Given the large sums spent on television advertising, even a small increase in an advertisement's persuasive effect might be profitable. Consider U.S. Super Bowl National Football League championship advertisements where, despite starting a year early, the creators violate many persuasion principles. (Readers can test that assertion by rating two or three Super Bowl commercials against the checklist.) There are exceptions, of course: Apple's "1984" commercial shows excellent compliance with principles.

(3) The effect of using recall as the objective criterion

Recall is widely regarded as important for the success of advertising, so it has face validity as an objective criterion. In our paper, we had found only one study with empirical evidence, and it showed a moderately strong relationship between recall and behavior.

In his commentary, Wright (this issue) argues that recall may in practice be superior to sales as an objective criterion given that sales data are confounded by influences other than advertising. Additional research using different objective criteria and better data would, nevertheless, be valuable.

(4) Representativeness of alternative methods tested

We consider the methods we tested to be broadly representative of the ways advertisers choose the most effective ads. Moreover, some of our reviewers are leading experts on copy testing, and one of the authors, Kesten Green, was the head of a market research firm. Of course, there are many alternative methods that might be used, and we did not attempt to identify and test them all. We hope that other researchers will conduct comparative tests of such methods. To aid their efforts, we fully disclose all methods and data in the AdPrin.com repository.

(5) Representativeness of research subjects

We were unable to obtain judgmental predictions from leading advertising experts. However, as we described in our paper (Armstrong et al., this issue), extensive

prior research in a variety of fields found that expertise in a domain has little effect on the accuracy of forecasts in complex situations. Moreover, Armstrong's (1980) review concluded that experts tend to be more over-confident than non-experts, resulting in greater resistance to adopting new findings. Given the number of studies to date on the topic, further research is unlikely to change those conclusions.

(6) Effect of combining on comparisons of methods

The PPI and PPIx methods for pretesting ads involve combining individual ratings by way of consensus. Combining is a well-established forecasting procedure. In our paper, we compared the accuracy of the PPI forecasts against the accuracy of individual judgmental forecasts, under the assumption that advertising experts are typically engaged individually.

Nevertheless, in response to S&H's suggestion that the relative success of the PPI might be due to combining—rather than the use of persuasion principles—we reanalyzed the data to compare combined unaided expert judgment forecasts with PPI forecasts. Larger advertisers might feasibly obtain the judgments of three unbiased experts, so we calculated all combinations of the forecasts of relative effectiveness of three unaided experts. In doing so, we distinguished between the self-reported experts recruited via Mechanical Turk, and the U.S. and China experts recruited by personal approaches.

The error reduction attributable to combining expert judgments averaged 17%, which is consistent with findings from prior research in different fields (Armstrong, Green and Graefe, 2015). Combining also reduced the errors of judgmental forecasts by 6% to 11% relative to forecasts from copy testing (see Table). Thus, one its own, combining three experts' forecasts offers a quick, simple and inexpensive way to obtain more accurate pre-test results. Contrary to S&H's conjecture, however, the error reductions from combining the unaided judgment forecasts of the U.S. and China experts' forecasts—the strongest of the alternative methods—provided only about one-third of the error reduction of the PPIx and less than a third of the error reduction of the PPI. Moreover, unlike the PPI, combining three experts' predictions does not provide advice on how to improve an advertisement.

Table: Accuracy of individual and combined forecasts

Method	N*	Forecast accuracy (%)		Error reduction‡
		Individual	Combined†	
Copy testing	288	n.a.	59.4	
Unaided judgment – novice	10,809	54.1	61.7	6
Unaided judgment – Mechanical Turk experts	2,444	54.9	63.1	9
Unaided judgment – U.S. and China experts	320	59.7	63.7	11
Theory-weighted PPI (PPIx)	480	61.7	72.4	32
Persuasion Principles Index (PPI)	480	61.0	74.5	**37**

* Number of individual forecasts of which of a pair of advertisements will be more effective.

† Forecast accuracy of combined *Copy testing* is the accuracy of the mode of the higher average intention-to-purchase response from the three procedures we used for each pair of advertisements. Forecast accuracy of combined *Unaided judgment – U.S. and China experts* is the average accuracy of all combinations of 3 experts' forecasts. Forecast accuracy of combined *Theory-weighted PPI (PPIx)* and *Persuasion Principles Index (PPI)* forecasts are the accuracy of the consensus of 3-or-more of 5 raters.

‡ Combined versus Copy testing.

Woodside: Checklists, Design Budget, PPI Difficulty, and Research Design

Woodside (this issue) provides additional support for checklists. A growing body of research has found that experts need *evidence-based* checklists to consistently make good decisions in complex situations. Checklists that are *not* evidence-based are common in management. For example, we are not aware of experimental evidence on relative profitability supporting Peters and Waterman's "eight attributes of management excellence" or Porter's "five forces" checklists. Such checklists are likely to harm decision-making by directing managers to do the wrong thing more efficiently.

Advertisers spend enormous amounts on media to spread their message. Why, Woodside asks, don't they invest substantially more on creating effective and profitable advertisements? Woodside cites Gross (1972), who showed that substantial gains could be achieved by creating alternative ads. That paper gained much attention among researchers, but got little coverage in textbooks and according to Irv Gross, little attention from practitioners. Gross's paper should receive greater attention now that practitioners can use the PPI to more accurately assess the most effective of a number of alternative ads.

Woodside argues that evaluating a pair of ads against the 195 persuasion principles included in the PPI "is hard work even with lots of practice." We understand that the task might look overwhelming to those who have not undertaken it. Hence, we designed the one-hour self-training module. After completing the training, our raters have found the task to be simple and quick.

Woodside wonders whether the PPI can be used to forecast *how* effective particular advertisements would be. We tested only whether the PPI could be used to predict the *more effective* ad, all else being equal, so we do not know the answer to that question. Nevertheless, the index method is well suited for developing quantitative forecasting models when sufficient data are available on the criterion variable. Model estimation would require a large number of ads with similar conditions—such as high involvement utilitarian products targeted to similar target markets. Regressing the index scores for the ads against the values of the criterion variable would provide the model coefficient (effect size) for those conditions. Such analyses might allow advertising research firms to develop and demonstrate expertise with some types of products, to the benefit of the firm and its clients.

Wright: Nature of the Tests and Recall, Persuasion, and Effectiveness

In his commentary, Wright (this issue) endorses our working definition of persuasion as "all influences…that lead people to action" and points out that we only examine the relative persuasiveness of print ads in isolation of other persuasive influences, such as media planning and competitor actions. We agree that further research testing the usefulness of the PPI when other aspects of persuasion are varied would provide important knowledge for advertisers.

Wright provides a useful discussion on why recall data provide a better measure of the persuasive effect of advertising than sales data. In practice, many influences other than ad exposure confound sales data to the extent that weak, but important effects of advertising cannot be detected. On the other hand, Wright argues, and we agree, that remembering an ad is important for persuasion. An unremembered ad is less likely to have an effect on sales.

Wright's discussion of memory theory might serve as a guide to test the effects of, and to propose new principles specifically used to improve a customer's memory about, a product. He references some existing principles on repetition (6.13), campaign contrast (8.3.1), and celebrity endorsements (6.6). Eleven other principles described in Armstrong (2010) relevant to memory are:

Attention: consistency across campaigns (8.2.1) and across time (8.2.2); slogans (8.4.1); brand identifiers (8.5.1 and 8.5.2); and color (8.7.1),

Activation: customer provides reason (5.11.1); customer asked to imagine

        satisfaction with product (5.11.2); and customer asked to make a

        prediction (5.11.4), and

Retrieval: customer asked to remember brand name or key arguments (5.11.5);

        and consider omitting or deleting key information (5.11.6).

Surely there are principles relevant to memory still to be identified or articulated. Moreover, further research on principles suggested by various advertising theories might also lead to the development of new evidence-based principles.

Discussion

The Persuasion Principles Index (PPI) can aid advertisers' creativity in two ways. First, each of the persuasion principles calls for creativity in their application. Second, given that the PPI can provide accurate predictions of relative effectiveness, advertisers benefit when they create more alternative ads to pretest.

In the past, practitioners could have been excused for missing useful knowledge about persuasion. It was spread across journals in different fields, tucked away in libraries, and written in arcane language. Now, novices can be trained in approximately an hour to use the body of knowledge about advertising persuasion, and they can then rate ads in less than half an hour using the PPI. The PPI checklists are freely available to practitioners at AdPrin.com. There are no patent restrictions, but we ask that users acknowledge the copyright. In references, we suggest referring to the date of the version that you use—the date on the first page of the software.

While our tests of the index method used print ads, TV commercials typically involve much larger expenditure for developing and running and so will often provide a more profitable application. Consider again the Super Bowl; advertisers invest enormous amounts for a chance to persuade the audience each year. Advertisers that employ the PPI to develop and choose the ad to run—especially those advertising high-involvement utilitarian products such as automobiles—could expect substantial benefits from doing so. The advertiser can easily calculate the PPI, and the cost to do so would be only a small fraction of one-percent of the budget.

The PPI can and should be added to the toolkit of advertising pretesting methods. In addition to having face and concurrent validity, our most recent study shows that the PPI scores higher on predictive validity than any of the other pretest methods examined. The procedure can be implemented using inexpensive services, such as Mechanical Turk,

and can be used with a rough mock-up of the ad, leading to operational ways to improve each ad tested.

Clients can use the PPI to evaluate advertisements proposed by agencies. More importantly, they can inform their agency that they will do so.

While our tests only examined advertisements for high-involvement products, the research behind the PPI applies to all persuasive messages and thus, to all ads. Nevertheless, the effect size—the extent of gain in persuasion arising from conformity to the persuasion principles in the PPI—will likely be smaller for many ads as outlined by Wright (this issue). For example, the effect size for ads of well-known, low-involvement products that contain no information would likely be much smaller.

The commentators indicated a desire for more knowledge on how best to predict and improve the effectiveness of advertising under diverse conditions. We agree. For example, we would like to see ongoing improvement in the persuasion principles and experiments on alternative implementations of the index method, comparative tests of alternative pretest methods, and comparative tests of alternative ways to communicate knowledge on advertising persuasion.

To encourage adoption of scientific findings on persuasion, we have added a free, self-directed course on Persuasive Advertising to AdPrin.com. The course includes a self-certification procedure along with a battery of tests. Universities and companies are welcome to use these materials to teach courses on persuasion principles.

Conclusions

Our main paper (Armstrong, Du, Green and Graefe, this issue) takes one of a series of steps that need to be taken to assess the validity of the persuasion principles. Specifically, we tested a prototype index model for assessing the extent to which an ad is consistent with the persuasion principles. Our findings suggest that ads that follow the principles *are* more effective.

While one might speculate that additional research would find different accuracy gains, or that better ways could be found to test for compliance with persuasion principles, these speculations should be regarded as hypotheses for further research, not as reasons to ignore immediate gains to be had by using the PPI.

Our analysis in this paper found that the evidence-weighted PPI index model procedure provided predictions that were more accurate than those from the theory-weighted PPIx index model specification.

Combining experts' unaided judgmental forecasts can stand on its own as a way to improve pretest accuracy. In particular, combinations of forecasts by three judges with modest expertise in advertising were about 10% more accurate than forecasts from the typical individual expert. The accuracy gain is one-third as large as the gain from using the PPI.

We hope that efforts to improve the cumulative body of knowledge on advertising will continue. The quest for knowledge on advertising must include efforts to test and refine existing principles, to develop and test new principles, and to develop and test procedures for using the Persuasion Principles Index model as a method for pretesting and improving advertisements.

References

Armstrong, J.S. (1980), "The seer-sucker theory: The value of experts in forecasting", *Technology Review*, Vol. 82 No. 7, pp. 16-24.

Armstrong, J.S. (1991), "Prediction of consumer behavior by experts and novices", *Journal of Consumer Research,* Vol. 18 No. 2, pp. 251-256.

Armstrong, J.S. (2010), *Persuasive Advertising: Evidence-based Principles*, Palgrave MacMillan, Hampshire, UK.

Armstrong, J.S. (2011), "Evidence-based advertising: An application to persuasion", *International Journal of Advertising*, Vol. 30 No. 5, pp. 743-767.

Armstrong, J.S. (2012), "Illusions in regression analysis", *International Journal of Forecasting*, Vol. 28 No. 3, pp. 689-694.

Armstrong, J.S., Du, R., Green, K.C. and Graefe, A. (2015), "Predictive validity of evidence-based persuasion principles", *European Journal of Marketing*, Vol. XX No. XX, pp. XXX-XXX.

Armstrong, J.S., Green, K.C. and Graefe, A. (2015), "Golden rule of forecasting: Be conservative", *Journal of Business Research,* Vol. 68 No. 8, pp. 1717-1731.

Armstrong, J.S. and Patnaik, S. (2009), "Using quasi-experimental data to develop empirical generalizations for persuasive advertising", *Journal of Advertising Research*, Vol. 49 No. 2, pp. 170-175.

Churchill, W. (1942), "The end of the beginning", The Lord Mayor's Luncheon, Mansion House, 10 November, available at: http://www.churchill-society-london.org.uk/EndoBegn.html (accessed 8 July 2015).

Graefe, A. (2015), "Improving forecasts using equally weighted predictors", *Journal of Business Research,* Vol. 68 No. 8, pp. 1792-1799.

Gross, I. (1972), "Creative aspects of advertising", *Sloan Management Review*, Vol. 14 No.1, pp. 83-109.

O'Keefe, D.J. (2015), "Evidence-based advertising using persuasion principles: Predictive validity and proof of concept", *European Journal of Marketing*, Vol. XX No. XX, pp. XXX-XXX.

O'Keefe, D.J. (2015b), "Message generalizations that support evidence-based persuasive message design: specifying the evidentiary requirements", *Health Communication*, Vol. 30 No. 2, pp. 106-113.

Sharp, B. and Hartnett, N. (2015), "Generalisability of advertising persuasion principles", *European Journal of Marketing*, Vol. XX No. XX, pp. XXX-XXX.

Woodside, A.G. (2015), "Predicting advertising execution effectiveness: Scale development and validation", *European Journal of Marketing*, Vol. XX No. XX, pp. XXX-XXX.

Wright, M. (2015), "Predicting what?: The strengths and limitations of a test of persuasive advertising principles", *European Journal of Marketing*, Vol. XX No. XX, pp. XXX-XXX.