Sawtooth Software

RESEARCH PAPER SERIES

The Number of Levels Effect in Conjoint: Where Does It Come From and Can It Be Eliminated?

Dick Wittink, Yale University
Joel Huber, Duke University
Peter Zandan, IntelliQuest
and
Richard M. Johnson, Sawtooth Software,
1992

530 W. Fir St.
Sequim, WA 98382
(360) 681-2300
www.sawtoothsoftware.com

THE NUMBER OF LEVELS EFFECT IN CONJOINT: WHERE DOES IT COME FROM, AND CAN IT BE ELIMINATED?

Dick R. Wittink
Cornell University
Joel Huber
Duke University
Peter Zandan
IntelliQuest, Inc.
Richard M. Johnson
Sawtooth Software
Published in Sawtooth Software Conference Proceedings, 1992

INTRODUCTION

In 1981, Currim *et al.* found, in a conjoint study with six attributes, three attributes defined on three levels and three defined on two levels, that the three-level attributes achieved much higher average importances than the two-level attributes. When they designed their study, they had not anticipated this result. Subsequently, in various experimental studies, such level effects have been demonstrated to exist for all popular data collection and parameter estimation methods.

In this paper, we define the attribute-level problem, we review results obtained in various experimental studies, and we discuss its relevance to commercial applications of conjoint analysis. Given a multitude of possible sources of this systematic effect, we motivate an experiment that is designed to provide further insight into the plausibility of alternative explanations for its existence. We interpret the experimental results, and we end with a discussion of how the level effect can be eliminated.

THE ATTRIBUTE-LEVEL PROBLEM

Currim *et al.* studied subscription series to performing arts events, using a variation of the tradeoff matrix data collection method (Johnson, 1974). Each tradeoff matrix involved one three-level and one two-level attribute. Pairs of objects were chosen from the matrices. Each pair involved a tradeoff. For example, one object would have the lower price, and the other object would have a superior benefit. Respondents were asked to indicate which of two objects they preferred for each of several pairs of objects. The strict preference data were used together with an assumed preference order for attribute levels (for example, the lowest price is most preferred) to derive complete rank orders for all the objects in a tradeoff matrix.

For managerial purposes, they summarized the results by computing average attribute importances. Importance was defined as the difference in partworths between the best and worst levels for an attribute. For each matrix involving one three-level and one two-level attribute, the importances for the two attributes summed to 100. In this study, the best and worst levels (for example, lowest and highest prices) also necessarily had the highest and lowest partworths. They found that the three-level attributes had average importances no less than 0.55 and no more than 0.66. The two-level attributes had average importances between 0.36 and 0.45. Computed separately for six segments, the three-level attributes' importances were no less than 0.49. The two-level attributes' importance never exceeded 0.50. Thus, attribute importances appeared to depend on the number of levels.

In subsequent studies, Wittink *et al. (1982*) and Wittink *et al. (1989*) showed that experimental manipulations of the number of attribute levels resulted in systematic differences in estimated attribute importances. Systematic level effects were found for full-profile rank orders, full-profile paired comparisons, full profile ratings, and tradeoff matrix ranks, using both metric (least-squares regression) and nonmetric (for example, LINMAP, in Srinivasan and Shocker, 1973) estimation methods. In one study involving five product categories, the relative importance of price was, on average, seven absolute percentage points higher when price had two additional intermediate levels (Wittink *et al.* 1989).

To investigate the research hypothesis that ACA (ACA System for Adaptive Conjoint Analysis by Sawtooth Software) might <u>not</u> suffer from a level effect, Wittink *et al. (1991*) compared results for ACA and full-profile ratings. The ACA results did include a level effect, although its magnitude was found to be only about half what occurred in the full-profile data. Their study, however, did not include experimental manipulations that might distinguish between different sources for the effect. For example, Green and Srinivasan (1990) suggest that the addition of intermediate levels for an attribute in the conjoint design may make a respondent pay more attention to that attribute. Currim *et al.* proposed a mathematical or algorithmic explanation.

<u>Managerial Relevance</u>. The existence of a level effect can have profound implications for the managerial conclusions obtained from a conjoint study. Consider for example, the results contained in Currim *et al.* One of their three-level attributes is a discount percentage. It was the least important of the three three-level attributes, but considering all six attributes it was the third most important. Currim *et al.* adjusted their results by considering the minimum and maximum possible importances for the attributes based on the number of levels in the conjoint design. After this adjustment, discount percentage was the least important of all six attributes. We expect managers to take different actions on the discount attribute when it is the least important as opposed to the third most important, in a list of six attributes.

For managerial purposes it is, nevertheless, much more pertinent to use market simulations than to rely on attribute importances. It is easy to demonstrate, however, that if attribute importances are affected by attribute levels, preference share predictions can be as well. Suppose, for example, that two attributes, A and B, can both have four levels. And imagine that in respondent X's mind the differences between the best and worst levels of both attributes are truly equally important. But if attribute A gets 4 and B gets 2 levels, A will obtain more relative importance. Conversely, if A gets 2 and B gets 4 levels, B will obtain higher importance.

To be precise, suppose that A will obtain a relative importance of 54 if it has 4 levels (and B would have a relative importance of 46 with 2 levels). But A will obtain a relative importance of 46 if it has 2 levels (so that B has an importance of 54 when it has 4 levels).

Now imagine that in a market simulation there are two products. Product I has the best level of A and the worst of B, while product II has the worst level of A but the best level of B. Then, based on the first-choice rule, respondent X would be predicted to choose product I if A has 4 (and B has 2) levels (because with 4 levels A is more important, and product I is superior on attribute A). But respondent X would be predicted to choose product II if A has 2 (and B has 4 levels).

If the choice rule involves a predicted probability of choosing an object, it is easier yet to demonstrate a level effect on market simulations. Since the importances are based on differences in partworths, it should be clear that subsequent computations that involve differences in or ratios of predicted utilities for objects will be systematically affected by the level effect.

THE LEVEL EFFECT IN ACA

Wittink *et al.* (1991) manipulated the number of levels for four attributes in a study of refrigerators. The attributes, Capacity, Energy Cost, Compressor Noise, and Price, had either two or four levels. For example, for Energy Cost the best and worst levels were $70 and $100, respectively. Half the respondents saw only those two levels, while the other half also saw two intermediate levels, $80 and $90. For full profile ratings, the relative importance of Energy Cost, based on the difference between the highest and lowest partworths (out of five attributes) was found to be 8.0 with two levels but 20.5 with four levels. For ACA, the relative importances were 13.7 for two and 19.9 for four levels of Energy Cost.

Across the four manipulated attributes, the difference in relative importance between the two- and four-level conditions was between 9 and 12 absolute percentage points for full profile, and between 5 and 6 absolute percentage points for ACA. Thus, the level effect was a much more serious phenomenon in the full-profile task than in ACA. But, because the ACA task proceeds in stages, Wittink *et al.* (1991) were able to trace the source of the level effect for ACA data.

They found no evidence of a level effect in the self-explicated data that provide the initial partworth solution in ACA. Thus, the level effect had to occur in the section in which preference intensity judgments are collected for one out of each pair of objects. Ideally, ACA chooses pairs of objects such that the predicted overall utility difference (based on the initial or updated partworth solution) between the objects is close to zero. An inspection of possible situations showed that if a respondent's self-explicated importances are incongruent (negatively correlated) with the numbers of levels assigned to the attributes for that respondent, it is impossible to achieve equality in the predicted utilities for objects defined on two attributes.

Wittink *et al.* reasoned that:

a) a bias in providing preference intensities toward the center of the scale can produce the levels effect;

b) such a bias is less likely to occur if the predicted utility differences are close to the center of the scale;

c) ACA is better able to produce pairs of objects with equal predicted utilities when respondents have higher importances for attributes with more levels.

This reasoning was used to construct the hypothesis that there should be a smaller levels effect for "congruent" respondents, that is, those for whom the more important attributes have more levels. Indeed, for respondents with congruent self-explicated importances and attribute levels, no level effect in ACA was found. On the other hand, for respondents with incongruent self-explicated importances and attribute levels, the magnitude of the level effect for ACA varied from 9 to 11 percentage points (or about twice the average level effect in ACA). This suggests that the magnitude of the level effect in ACA depends on the difference in predicted utility between two objects. Thus, the results in Wittink *et al.* (1991) <u>suggest</u> an algorithmic explanation for the level effect. Yet, alternative explanations cannot be ruled out.

EXPERIMENTAL DESIGN

To obtain further insight into the plausibility of alternative explanations we conducted another study with four manipulations: (1) the number of attribute levels (for four attributes); (2) the balancing of paired objects in terms of predicted utilities; (3) a tutorial on the meaning of the attributes; and (4) the inclusion of prompts in the preference intensity section of ACA.

Based on the consistent results obtained in the earlier studies referred to, we expect an attribute's importance to be higher with more levels. The second manipulation (balancing) is expected to influence the magnitude of the level effect. The closer the predicted utilities of the objects, the smaller the <u>potential</u> for a level effect.

The third and fourth manipulations represent attempts to study potential behavioral explanations of the effect. The third manipulation consisted of making tutorial information available to half the respondents. Respondents who do not fully understand the attributes may react to cues unintentionally provided by the researcher. For example, respondents may infer that attributes with more levels should have more importance. If so, we would expect additional information about the attributes to increase respondents' understanding, such that the results better reflect their "true" importances. In this manner, the level effect may be reduced under the tutorial treatment.

The fourth manipulation, "prompting," was intended to jar respondents into a higher state of awareness. A bored respondent may provide answers with a substantially random component and may have a larger midscale bias. Thus, we hypothesized that an occasional challenge to an answer provided by the respondent might produce better responses. At randomly chosen points in the interview the screen turned red with the message: "Are you sure about your answer of x? Please think some more about the strength of your preference. Press any key now to answer the question again."

The authors' expectations differed for this experiment. Huber, Johnson, and Zandan favored a "behavioral" explanation for the level effect, while Wittink favored an "algorithmic" one. A behavioral source had been illustrated in a study reported by Johnson (1992). Respondents were asked the dollar values of improvements in TV sets. Values of improvements from the worst to the best level of each attribute were found to be greater if attributes had intermediate levels. Since importance was stated directly by the respondent, rather than estimated by an algorithm, that study suggested a behavioral origin of the effect.

The algorithmic argument, on the other hand, concerned the ability of ACA to produce pairs of objects with nearly equal estimated utilities for each respondent. ACA ordinarily tries to produce pairs that are "balanced" in the sense of both objects having nearly equal utility. Although we couldn't produce a version of ACA that did a better than usual job of this, we were able to modify ACA to do a worse job, by disabling the part of the program that does the balancing. The algorithmic hypothesis would suggest that respondents who received balanced pairs would display a smaller level effect than those for whom the pairs were not balanced.

The experimental design to test hypotheses about behavioral and algorithmic explanations of the level effect was a 24 full factorial design. The product category chosen was the notebook computer. Six attributes were used: (1) brand name; (2) notebook size; (3) weight; (4) battery life; (5) performance; and (6) purchase price. Half the respondents saw two levels of notebook size and battery life but four levels of weight and purchase price. The allocation of these levels to the attributes was reversed for the other half of the respondents.

RESULTS

Data were obtained from 403 respondents (40 percent) out of 1,008 surveys mailed. The sample was drawn from the office intensive file of Dun and Bradstreet and screened over the telephone by the market research firm of IntelliQuest to locate an individual with responsibility for purchasing or using notebook computers.

For each respondent, attribute importances were calculated based on the difference between the largest and smallest partworths for each attribute. Relative importances (by making the importances sum to 100) were analyzed as a function of all main and interaction effects for the experimental manipulations. The relative importance equations of the two attributes for which no level manipulation occurred had no significant explanatory power. For the other four equations, the level effect was significant ($p < .01$) in each case. And the level effect was significantly higher without balance than with balance for two of the attributes. No other consistent effects were obtained.

The interaction effect between level and balance was negative for all four attributes. Thus, in all four cases did the absence of balance in ACA increase the level effect. This effect cannot be attributed to a behavioral phenomenon involving the level manipulation. And, because neither of the two manipulations designed to produce a behavioral effect showed a significant interaction effect, we conclude that the evidence from this study is entirely consistent with an algorithmic explanation.

The Balance Manipulation. ACA includes a section that eliminates dominated pairs of objects from consideration. For example, if Price and Battery Life are the two attributes, the commercial version of ACA would not select one object to be better than the other on both Price and Battery Life. The selection of only nondominated pairs is what we call the existence of balance.

By eliminating this section from ACA for half the respondents, we allowed the pairs of objects to be both nondominated and dominated. The result is that the expected difference in predicted utilities for paired objects is now further from zero. If respondents have a tendency to provide preference intensity ratings toward the midpoint of the scale, the opportunity for distortion is greater when dominated pairs of objects are allowed.

It turns out, however, that the nature of the possible interaction between the level and balance manipulation is very complex. We will address this issue in detail in a future paper. We restrict ourselves here to the empirical results.

We show the average relative importances for all respondents, and separately for the combinations of levels and balance, in Table 1. Although we show separate averages for all four attributes, only the difference (in average difference between 2 and 4 levels) between the balance alternatives for Size and Price are significant ($p < .01$). Considering all four attributes, the level effect is between 1.34 and 4.00 when balance exists, but it is between 4.60 and 6.76 in the absence of balance. Thus, the commercial version of ACA with nondominated pairs is favored over a version that also allows for dominated pairs of objects.

Table 1

Average Relative Importances, Overall and for

Level and Balance Combinations

		Experimental Manipulations			
	Balance	Yes		No	
Attribute	**Overall Average** / Levels	**2**	**4**	**2**	**4**
Size	9.52	7.25	11.25	6.43	13.19
		4.00		6.76	
Weight	12.05	10.22	13.78	9.80	14.40
		3.56		4.60	
Life	15.97	14.48	17.04	13.74	18.50
		2.56		4.76	
Price	22.06	22.42	23.76	18.14	23.92
		1.34		5.78	
Average Difference Between 2 and 4 Levels		2.84		5.48	

Noise-Adjusted Importance Measure. Even though the number of levels was manipulated for attributes that can be expected to have a monotone preference function, it is possible for the partworths to violate monotonicity. And, of course, violations of monotonicity are more likely to happen for attributes with a larger number of levels. Thus, it is possible that the magnitude of a level effect is reduced when we use an importance measure that is not influenced by statistical noise.

An alternative importance measure is to define importance based on the difference in partworths for the extreme levels. All four attributes with level manipulations in our study are assumed to be monotonically related to preference. That is, the lower the price, other things being equal, the more preferred an object. Similarly, the smaller the size, the lower the weight, and the longer the battery life, the more attractive a notebook is expected to be. By defining the importance using the difference in partworths for the extreme levels, we also ensure consistency across the level manipulations. For example, this measure now captures the importance of the difference between \$3,600 and \$2,300 (the extreme prices) for Price, in all cases (as is also true for the self-explicated importance in ACA). However, we note that in the first paper on level effects (Currim *et al.)* statistical noise played no role. In that paper, the highest and lowest partworths necessarily occurred for the best and worst levels defined *a priori.*

We estimated the effects of the experimental manipulations on this noise-adjusted measure of importance in exactly the same manner as for the original importance measure. Significant level effects occurred for Size ($p < .01$), Life ($p < .01$), and Price ($p < .05$), but not for Weight. The balance manipulation produced a significant interaction effect with the attribute levels only for Price ($p < .05$). However, with the exception of Weight, the interaction effect has the expected sign. Neither the prompting nor the tutorial manipulations showed significant interactions with the level manipulation. For comparison purposes, we show the average relative (noise-adjusted) importances for all respondents, and separately for the combinations of levels and balance, for all four attributes in Table 2.

The overall average importances do not appear to be very different when we compare the first column in Table 2 with the first one in Table 1. The importances of Life and Price are now somewhat higher, implying that one or both of the other attributes (Brand Name and Performance) have lower importances. The level effect in the balance condition is now between -0.10 and 2.24, while in the other condition it is between 0.68 and 4.24. On average, the level effect is approximately half in Table 2 of what it is in Table 1.

Table 2

Average Relative Noise-Adjusted Importances, Overall and for

Level and Balance Combinations

		Experimental Manipulations			
	Balance	Yes		No	
	Levels	2	4	2	4
Attribute	**Overall Average**				
Size	9.33	8.24	9.92	7.68	11.48
		1.68		3.80	
Weight	12.70	12.33	13.73	12.03	12.71
		1.40		0.68	
Life	17.99	16.73	18.97	16.49	19.77
		2.24		3.28	
Price	25.80	27.06	26.96	22.42	26.76
		-0.10		4.24	
Average Difference Between 2 and 4 Levels		1.31		3.00	

DISCUSSION

The results of our study indicate that if ACA did not exclude dominated pairs of objects from consideration, the level effect would be larger (at least in this application). On average, the difference in importance when an attribute has four versus two levels is 2.84 absolute percentage points for the current version of ACA but 5.48 absolute percentage points when dominated objects are allowed to occur in the preference intensity section. The noise-adjusted importance measure shows that the level effect is reduced, but not eliminated.

Our study did not provide statistical support for the manipulations designed to capture behavioral effects. This does not mean that respondents do not react to the number of levels included for a given attribute. All we can claim is that the prompt and tutorial manipulations in this application were insufficient. It is possible that other manipulations can provide a substantial reduction in the level effect. The results of this study suggest that the source of the effect is algorithmic.

Given that the balance manipulation reduced the level effect to about half its magnitude otherwise, it is appropriate to ask whether ACA can be improved further. Wittink *et al.* (1991) have shown that the level effect, in a study of refrigerators, was zero for respondents with congruent attribute levels and self-explicated importances. This means that if a respondent indicates that attribute A is very important and B is not important, A should have, say, 4 and B have 2 levels. In that case, the predicted utilities for objects defined on attributes A and B, selected by ACA, are approximately equal. Thus, if ACA is modified to increase the likelihood of selecting pairs of objects with equal predicted utilities, the level effect will become smaller.

The more important an attribute is, the more reason we have to include intermediate levels, both to avoid level effects and to maximize useful learning. That is, there is no need to learn the partworths for intermediate levels if an attribute is unimportant. Thus, the more important an attribute is for a given respondent, the more reason we have to include intermediate levels in order to understand the shape of the function and to minimize the level effects. Our results suggest that the next ACA version would benefit if modified to reflect this idea. In the meantime, conjoint analysts can improve the results by using a larger number of levels for the attributes with higher (expected) importances. This information can be obtained in a pre-test of respondents (or in a procedure such as the self-explicated portion in ACA).

Finally, we want to emphasize that despite the absence of evidence in favor of psychological explanations of the level effect, it seems to us that it is always a good idea to make respondents as motivated and as smart as possible.

We thank Karlan Witt of IntelliOuest for her assistance with various aspects of the study.

REFERENCES

Currim, Imran S., Charles B. Weinberg, and Dick R. Wittink (1981). “The Design of Subscription Programs for a Performing Arts Series.”*Journal of Consumer Research,* 8 (June), 67-75.

Green, Paul E. and V. Srinivasan (1990). “Conjoint Analysis in Marketing Research: New Developments and Directions.” *Journal of Marketing,* 54 (October), 3-19.

Johnson, Richard M. (1974). “Trade-Off Analysis of Consumer Values.” *Journal of Marketing Research,* 11 (May), 121 -7.

Johnson, Richard M. (1992). “Comment on Attribute Level Effects.” *Second Annual Advanced Research Techniques Forum Proceedings,* American Marketing Association, 62-4.

Srinivasan, V. and Allan D. Shocker (1973). “Linear Programming Techniques for Multidimensional Analysis of Preferences.” *Psychometrika,* 38, 337-69.

Wittink, Dick R., Lakshman Krishnamurthi, and Julia B. Nutter (1982). “Comparing Derived Importance Weights Across Attributes.” *Journal of Consumer Research,* 8 (March), 471 -4.

Wittink, Dick R., Lakshman Krishnamurthi, and David J. Reibstein, (1989). “The Effects of Differences in the Number of Attribute Levels on Conjoint Results.” *Marketing Letters,* (2), 113-23.

Wittink, Dick R., Joel C. Huber, John A. Fiedler, and Richard L. Miller (1991). “The Magnitude of and an Explanationl Solution for the Number of Levels Effect in Conjoint Analysis.” working paper, November.