Available online at www.sciencedirect.com
SCIENCE @ DIRECT®
Social Science Research 35 (2006) 279–294
Social Science RESEARCH
www.elsevier.com/locate/ssresearch

Additive and multiplicative effects in a fixed 2×2 design using ANOVA can be difficult to differentiate: Demonstration and mathematical reasons ☆

Johannes A. Landsheer [a,*], Godfried van den Wittenboer [b], Gerard H. Maassen [a]

[a] *Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht University P.O. Box 80140, NL3508TC, The Netherlands*
[b] *Department of Education, University of Amsterdam, The Netherlands*

Available online 22 January 2005

Abstract

Several researchers have expressed doubts as to whether ANOVA is fit to detect interaction in a fixed 2×2 design. To study this, values of a dependent variable are created with linear construction formulae, with and without a product term. Our simulation demonstrates that when the dependent variable results from a product term, ANOVA often fails to detect a significant interaction effect, while at the same time two strong additive main effects are indicated. As a consequence, multiplicative models are rejected in favor of an additive model. Mathematically it is shown that the ANOVA results can indicate strong main effects as the result of a solely multiplicative contribution of both factors. It is recommended to abandon the fixed ANOVA model for the study of interaction effects and use a regression model with assessed level values of each independent variable instead.

☆ We thank our colleague Olav Laudy who contributed greatly by programming the Mathcad simulation.
* Corresponding author. Fax: +31 30 2535797.
E-mail address: j.a.Landsheer@fss.uu.nl (J.A. Landsheer).
URL: http://www.fss.uu.nl/ms/hl/anovasim.htm. (J.A. Landsheer)

0049-089X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ssresearch.2004.10.004

1. Introduction

Since the publication of *The Design of Experiments* by Fisher (1935), the analysis of variance has become one of the most important statistical techniques for the study of treatment effects. However, the interpretation of combinatory or multiplicative effects, and its differentiation from additive effects has been the subject of concern. Hays (1973) wrote that the subject of interaction is neither elementary nor fully explored. Since then, various authors have restated this viewpoint. It is seldom a straightforward matter to interpret the combination of main effects and interaction, even in complete factorial designs (Jaccard, 1998). Rosnow and Rosenthal (1995) discussed the importance of evaluating significant main effects, as well as significant interaction effects and warned against the idea that main effects can be disregarded as meaningless once an interaction effect has been obtained. Unbalanced designs can create even larger interpretation problems (Searle, 1987). Beside problems of interpretation, there is considerable doubt as to whether ANOVA is sufficiently sensitive to detect interaction effects when these are present in the data. Wahlsten (1991) identified the problem that the sample size required for the detection of interaction effects is seven to nine times larger than that needed to detect main effects. In comparison with the main effects tests, McClelland and Judd (1993) estimated that interaction tests are 80% less efficient, but unlike Wahlsten they attribute this reduction mainly to the use of field studies, whilst claiming that experimental studies are not so affected. Jaccard and Wan (1995) argued that, in non-experimental designs, the unreliability of measures might affect the detection of interaction effects, because the reliability of product terms is often quite low compared with the reliability of its components.

The relevance of a possible underestimation of interaction effects should itself not be underrated. In both experimental and non-experimental research, ignoring interaction effects has a direct influence on theoretical developments. If interaction effects are underestimated, this leads to the adoption of simple additive models when, in fact, models should have been developed to take the combinatorial effects of independent variables into account. For instance, the fact that analysis of variance often fails to detect interaction effects led Wahlsten (1990) to argue that this distorts the heredity-environment discussion, because it makes it difficult to estimate the unique contribution of a 'heredity × environment' interaction effect compared with the main effects.

The interaction of AB concerns the effect above and beyond what can be accounted for by addition of the effects of factors A and B. In other words, the interaction is the partialled product of A and B (Cohen, 1978; Cohen and Cohen, 1983). There are different approaches to describe interaction, but mathematically they are equivalent and should not lead to contradictory interpretations.

The most common way to describe a model with interaction is $Y = p*A + q*B + r*A*B + s$, where the combination $A*B$ represents the interaction. As identified by Cohen and Cohen (1983), whether we are dealing with nominal or quantitative independent variables, interactions are carried by products of variables. In the additive model $Y = p*A + q*B + s$, and the dependent variable is affected in such a way that a change in the dependent variable Y as a consequence of A is identical for each level of B. Usually, the additive model is considered fitting when the interaction is not significant.

Another approach is to look at the interaction as the effect of one of the factors which is moderated by the other factor: $Y = p*A + B*(q + r*A) + s$, or $Y = q*B + A*(p + r*B) + s$. The choice as to which of the factors is the moderator and which the moderated variable, has to be decided on theoretical grounds (Jaccard and Guilamo-Ramos, 2002). Saunders (1956) suggested this approach; both McClelland and Judd (1993) and Hayduk and Wonnacut (1980) provide interesting discussions of this model.

Confining ourselves to 2×2 designs with fixed factors, we will demonstrate that ANOVA or regression analysis applied to data resulting from a multiplicative model often indicate only strong main effects, even when a balanced design is used. Our simulation shows that a product term cannot always be detected, since the interaction effect does not show up under certain circumstances, while at the same time relatively large and therefore often significant main effects occur. As a result, an analysis of variance will often indicate that the multiplicative model should be rejected in favor of an additive model. We discuss the mathematical background of our simulation and we argue that this model is often applicable. Further consequences and strategies to cope with this phenomenon are discussed in the concluding section.

2. Method

In a fixed 2×2 design, two factors are used, each consisting of two qualitatively different levels or groups. These qualitative differences are either the result of the experimental treatment or the result of classification and the groups are identified by means of arbitrary chosen codes. Treatment differences as such are not assessed, instead the differences on the dependent variable are used to measure effects and the independent variables A and B represent treated groups, not treatment strengths or specific realized results of each treatment. If groups are assembled by blocking, the categories can often also be considered as an indirect representation of the true variable. For instance, the differences between man and woman can be represented more precisely by variables such as hormonal levels, length or strength, which may influence the dependent variable more directly.

When direct measurement of the relevant variables is not possible, it makes sense to consider the independent variables A and B as indirect representations of the true variables A' and B'. As an example, we use a model of Vroom (1964): performance is a function of Ability × Motivation. These variables are not directly observable. Let us suppose an experiment, where both Ability and Motivation are manipulated, resulting in four groups. Furthermore, we suppose the relationship $Y = A' * B'$, where Y is the dependent variable Performance, and A' and B' represent the true variables Ability and Motivation. In ANOVA, both variables are represented by the nominal independent variables A and B, both indicating different groups.

Now, an observed dependent variable Y can be created on the basis of the assumed relationship between Y and the variables A' and B', where the values of A' and B' are the respective true result of each administered treatment. In general a linear dependent variable can be constructed with a formula: $Y = p * A' + q * B' + r * A' * B' + s + e$, where p, q, r and s are constant weights and e represents the normally distributed residual error component. To study the conditions for detecting combinatorial effects, we create values of a dependent variable Y with such construction formulae. In this way, the relation between true independent variables and the observed dependent variable has been defined.

In this paper, we confine ourselves to a balanced 2×2 experiment, in which neither the independent variables A and B, nor true variables A' and B' are correlated. Since ANOVA uses contrast coding of the factor levels as a default in a fixed 2×2 design, the code values of A and B add up to zero. This is convenient when studying interaction effects, because it prevents possible covariance between each of the independent variables and their product. For 2×2 designs, ANOVA leads, consequently, to an orthogonal analysis model, in which all correlations between the contrasts of the main vectors A, B and the interaction vector $A * B$ are zero. Orthogonal coding is commonly used in computer programs for the fixed model analysis of variance. When non-contrasting codes would be used in the analysis, this interaction vector is generally correlated with the main effect vectors (Bohrstedt and Goldberger, 1969; McClelland and Judd, 1993). This led Cohen and Cohen (1983) to the explicit use of contrasting orthogonal codes for the levels of the independent variables when using multiple regression. In this manner, analysis with multiple regression leads to results equivalent to those of ANOVA and to the same conclusions.

Regardless of the program used, ANOVA with fixed factors translates the values of the levels into -1 and $+1$ in case of 2×2 designs. In most ANOVA-programs these codes are shown when the researcher requests the coefficient contrast matrix. In all cases, main effects are carried by the variables A and B, while the interaction vector is computed as the cell wise product of A and B. ANOVA is commonly considered to be an appropriate technique to detect an interaction effect, and hence to differentiate between multiplicative and additive models.

The question to be studied is under what conditions ANOVA will be able to indicate the multiplicative effects introduced in the data as significant. First, we will demonstrate how often an interaction effect is rejected. Next, we review mathematically a number of construction formulas to generate data, varying from purely multiplicative to purely additive.

3. A demonstration

Product term analysis will be demonstrated by means of an artificial experiment of the performance model of Vroom (1964): Performance = Ability × Motivation. In this example, two levels (high and low) are chosen for ability (A) and for motivation (B), so that a simple 2×2 design will be obtained (see the SPSS code in the Appendix). Both the high level of ability and the high level of motivation are the result of manipulation. As the number of observations is equal in each cell, the design is balanced.

To calculate the performance Y using the formula $Y = A' * B'$ in this example, we have assumed a low ability level of 2 for subjects who do not receive treatment to increase their ability, while the subjects in the high ability condition acquire a heightened ability level of 4 as a result of the manipulation. Therefore, we assume that there is a certain positive ability in untreated subjects, while treated subjects acquire a higher ability. Similarly for motivation B', we assume a low positive motivation of 2 in untreated subjects and a higher value of 4 for the subjects who have received the treatment to increase their motivation. Of course, when measuring Y we may expect some error to occur, and therefore some normally distributed error has been added as well. Our demonstration concerns a sample of 96 cases divided over the 4 cells of the 2×2 design with 24 cases in each cell. The SPSS setup is shown in the Appendix. Fig. 1 shows an example of the mean scores.

The ANOVA results in Table 1 show two strong main effects for A and B, but a failing interaction effect. Only if a much larger sample had been used, would ANOVA have been able to detect a small, but significant interaction. Although the data are created with a combinatory effect of ability and motivation, the results of ANOVA are difficult to interpret as a product relation $A * B$. Instead, the significant main effects draws the attention to an additive model.

As additive and multiplicative models concerns the relation between independent and dependent variables, it may be useful to use multiple regression (MR) as a general variance-accounting procedure (Cohen, 1968; Cohen and Cohen, 1983). This offers the b values as a direct representation of the relationship between independent and dependent variables. Table 2 shows the results of the same data used in Table 1. The regression analysis results show that the b-coefficient of the interaction effect is not significant. Furthermore, the R^2-change statistics of the comparison of the model with and without interaction, shows that the model with interaction does not significantly contribute to the explanation of variance. In short, the MR-results are similar to the results of ANOVA and they also lead to the failure of rejecting H_0 for an interaction effect. According to both methods, an additive model would have prevailed over a multiplicative model, even though all values of the dependent variable resulted from a multiplicative relationship.

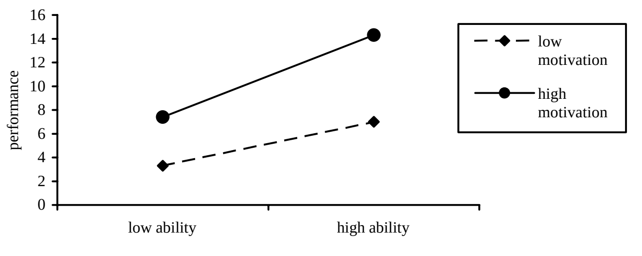

Fig. 1. Means of the non-additive relation in the 2×2 design of Tables 1 and 2 that are produced by a single product term. ANOVA analysis does not show a significant interaction effect ($N = 96$).

Table 1
ANOVA results for a reconstruction of Vroom's (1964) product term theory Performance = Ability × Motivation

Source	Sum of squares	*df*	Mean squares	*F*	*p*
Intercept	6176.2	1	6176.2	116.0	.00
A	679.4	1	679.4	12.8	.00
B	779.6	1	779.6	14.6	.00
$A \times B$	57.1	1	57.1	1.1	.30
Error	4899.0	92	53.2		
Total	12591.3	96			

Variable *A* represents with two levels the true treatment strength A' for ability, and variable *B* represents with two levels the true treatment strength for motivation (B'). The yield (performance *Y*) is calculated as the result of the construction formula $Y = A' * B' + e$, where it is assumed that the realized mean values on both variables A' and B' are 2 and 4.

Table 2
Multiple regression results for a reconstruction of Vroom's (1964) product term theory Performance = Ability × Motivation. Both *A* and *B* are coded orthogonally

Model		*B*	*SE*	*β*	*t*	*p*	Model R^2 change	Significance
$Y = A + B$	Constant	8.0	.8		10.8	.00	0.12	.00
	A	2.7	.7	.3	3.6	.00		
	B	2.9	.7	.4	3.8	.00		
$Y = A + B + A * B$	Constant	8.0	.7		10.8	.00	.01	.30
	A	2.7	.7	.3	3.6	.00		
	B	2.9	.7	.4	3.8	.00		
	AB	.8	.7	.1	1.0	.10		

The yield (performance *Y*) is calculated as the result of the construction formula $Y = A' * B' + e$, where it is assumed that the realized mean values on both variables A' and B' are 2 and 4.

To demonstrate the seriousness of the problem an extended simulation was created in Mathcad. Fig. 2 shows the simulation results for a total of 21.000 runs. In all cases the performance results from the Performance = Ability * Motivation relationship. Apart from the error component, we assumed a motivation of 1 for the untreated 'low motivation' subjects and an ability of 1 for the 'low ability' subjects. The ability of the 'high ability' subjects is assumed to be constant at a level of 1.5. In this simulation the manipulation results of motivation is varied from 1 to 3 in 21 steps, while the ability of the subjects who received a treatment to increase their ability is assumed to be constant (1.5). For each increase of motivation 1000 ANOVA's are calculated, each with a random error component. The error component is normally distributed with a mean of zero and a standard deviation of 1. Fig. 2 plots the proportion of statistical significant results ($p < .05$) for main effect motivation, main effect ability and the interaction effect ability * motivation.

The proportion of significant effects of ability grows from 60 to 100%, while in fact the manipulation difference (apart from the error component) is assumed to be constant at .5. The proportion significant main effects of motivation rapidly grows

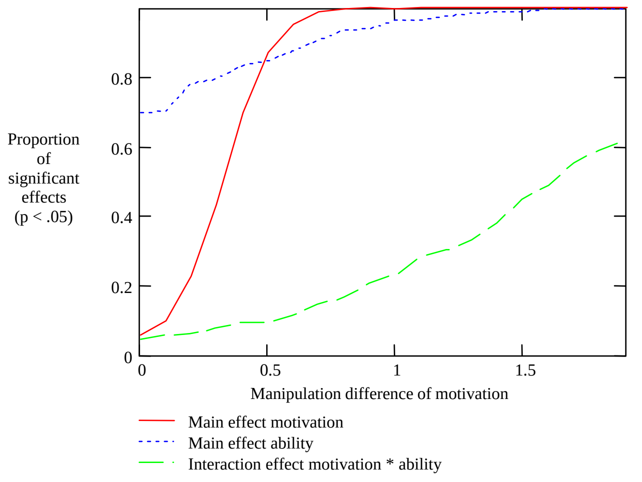

Fig. 2. ANOVA results of 21.000 simulations of a combinatory effect Performance = Ability × Motivation. The mean differences between the groups are kept constant (.5) for ability, while the mean manipulation difference of motivation is varied in 21 steps from 0 to 2. For each step the simulation is repeated 1000 times.

to 100%, with the simulated motivation difference growing from 0 to about .75. In contrast, the proportion of significant interactions of ability ∗ motivation slowly grows form .05 to about 60% when the maximum motivation difference of 2 between both groups has been reached. The proportion of significant interaction effects remains clearly behind that of the main effects. This result is unexpected as the relationship between the factors is purely multiplicative.

4. Mathematical models

The problem of significant main effects and a product term that fails to be significant is not confined to these simulations. It stems from the analysis scheme of ANOVA when product terms constitute the data. Suppose, for example, that both independent variables can vary freely and that any effect size may result can result from their influence. If the cell means are O_1, O_2, O_3, and O_4, and the 2×2 design is balanced with n observations in each cell, the total mean will be

$$X_{\text{meantot}} = (O_1 + O_2 + O_3 + O_4)/4$$

and the Sums of Squares will be

$$SS_A = n(O_1 + O_2 - O_3 - O_4)^2/4,$$

$$SS_B = n(O_1 - O_2 + O_3 - O_4)^2/4,$$

$$SS_{AB} = n(O_1 - O_2 - O_3 + O_4)^2/4.$$

4.1. A multiplicative model

Suppose now that the construction equation creating the values of the dependent variable consists of a single product term component. The relationship between the independent variables and the dependent variables is then simply, $Y = A' * B' + e$ with e as the error term. Let the levels of A' be a_1 and a_2, the levels of B' be b_1 and b_2, and the mean error in cell i be m_i. Consequently, the outcome in each of the four cells will be: $O_1 = a_1 \times b_1 + m_1$, $O_2 = a_1 \times b_2 + m_2$, $O_3 = a_2 \times b_1 + m_3$, and $O_4 = a_2 \times b_2 + m_4$. Substituting these values in SS_{A}, SS_{B}, and SS_{AB} we obtain:

$$SS_A = n(a_1 \times b_1 + m_1 + a_1 \times b_2 + m_2 - a_2 \times b_1 - m_3 - a_2 \times b_2 - m_4)^2/4,$$

$$SS_B = n(a_1 \times b_1 + m_1 - a_1 \times b_2 - m_2 + a_2 \times b_1 + m_3 - sa_2 \times b_2 - m_4)^2/4,$$

$$SS_{AB} = n(a_1 \times b_1 + m_1 - a_1 \times b_2 - m_2 - a_2 \times b_1 - m_3 + a_2 \times b_2 + m_4)^2/4.$$

Since the mutually independent error means m_{i} are uncorrelated with the factors A' and B', and because their variances equal $E[m_i^2] = \sigma^2/n$, the expected mean squares yield

$$E[MS_A] = [n\{(b_1 + b_2)(a_1 - a_2)\}^2/4] + \sigma^2, \tag{1}$$

$$E[MS_B] = [n\{(a_1 + a_2)(b_1 - b_2)\}^2/4] + \sigma^2, \tag{2}$$

$$E[MS_{AB}] = [n\{(a_1 - a_2)(b_1 - b_2)\}^2/4] + \sigma^2, \tag{3}$$

while

$$E[MS_{\mathrm{within}}] = \sigma^2.$$

If the null hypothesis of no effect of main factors A or B had been true, $E[MS_A]$ and $E[MS_B]$ would have been σ^2. Apart from the value of n and σ^2, the value of $E[MS_A]$ will depend on the difference $a_1 - a_2$ for a main effect of A', but also on the sum $b_1 + b_2$. Even if the difference $a_1 - a_2$ is small, it might be outweighed by the sum $b_1 + b_2$. The same holds true for the effect of main factor B', but this effect depends on the difference $b_1 - b_2$ and the sum $a_1 + a_2$, respectively. Neither $a_1 + a_2$ nor $b_1 + b_2$ reflect treatment differences and make the main effects hard to interpret. Only if the other variable consists of contrasting values for the levels ($a_1 = -a_2$ and $b_1 = -b_2$) do these main effects become zero, because $b_1 + b_2 = 0$ and $a_1 + a_2 = 0$.

Main effects will be significant in a 2×2 design with $df = 1$ for each effect in the numerator and $df = 4(n - 1) = N - 4$ for the denominator, if

$$F = \frac{MS_{\text{effect}}}{MS_{\text{within}}} \geqslant c,$$

where c is the critical level of F given α and the degrees of freedom. In other words, the F value of such a main effect depends on the values of the other variable: if $b_1 + b_2$ or $a_1 + a_2$ increase, the main effect of, respectively, A and B is bound to increase. In other words: in the case of a multiplicative effect, non-orthogonal treatment values influence the main effects.

This demonstrates that in the case of solely a combined effect of both variables, main effects are to be expected next to an interaction. Since the expected interaction mean square $E[MS_{AB}]$ depends on the product of the difference $a_1 - a_2$ and $b_1 - b_2$ at both variables, only these differences will be of influence on $E[MS_{AB}]$. Only the treatment differences contribute to this effect and the position of the zero point is of no relevance.

4.2. Composite multiplicative model

Suppose again that the mean level values of A' are a_1 and a_2, and of B', b_1 and b_2. Furthermore, suppose that the effects depend directly on the value of the independent variables and that main effects are present next to the product term, that is: $Y = A' + B' + A' * B' + e = (A' + 1) * (B' + 1) - 1 + e$. Again, let cell means be O_1, O_2, O_3, and O_4. Now, O_1 becomes $O_1 = a_1 + b_1 + a_1 \times b_1 + m_1$, and so on. Consequently, the expected mean squares yield

$$E[MS_A] = [n\{(a_1 - a_2)(b_1 + b_2 + 2)\}^2/4] + \sigma^2, \tag{4}$$

$$E[MS_B] = [n\{(b_1 - b_2)(a_1 + a_2 + 2)\}^2/4] + \sigma^2, \tag{5}$$

$$E[MS_{\text{AB}}] = [n\{(a_1 - a_2)(b_1 - b_2)\}^2/4] + \sigma^2. \tag{6}$$

Again, both $E[MS_A]$ and $E[MS_B]$ will be dependent on both independent variables. Besides the effects of A' and B' themselves, quite naturally thought off as the treatment differences $a_1 - a_2$ and $b_1 - b_2$, both expected mean squares are dependent again on the added values of the other variable. In addition, the expected mean square $E[MS_{AB}]$ in itself does not depend on the sum of the values of either of the variables. It depends on the product of both level differences and represents the combinatorial effect of both variables. In fact, it is equal to Eq. (3) of the multiplicative model.

4.3. Additive model

Things become rather different when only main effects are involved. The construction equation is $Y = A' + B' + e$ and the expected mean squares yield

$$E[MS_A] = \{n(2a_1 - 2a_2)^2/4\}^2 + \sigma^2, \quad (7)$$

$$E[MS_B] = \{n(2b_1 - 2b_2)^2/4\} + \sigma^2, \quad (8)$$

$$E[MS_{AB}] = \sigma^2. \quad (9)$$

Neglecting n and σ^2 again, ANOVA now delivers main effects that are solely dependent on the difference between $a_1 - a_2$, and $b_1 - b_2$. There is no product term contribution, as might be expected. It is also clear that both main effects are completely independent of each other. In other words: both independent variables can be positioned freely, and the main effect of a factor is determined only by the treatment differences.

4.4. Main effect dependencies

Fig. 3 illustrates the dependency of MS_A on the treatment values b_1 and b_2, if no error term is included in the construction equation. The treatment differences $a_1 - a_2$ and $b_1 - b_2$ are kept at the same level (2), but b_1 is moved on the scale from -5 to 5, and hence b_2 from -3 to 7 and $b_1 + b_2$ varies from -8 to 12. The interaction MS_{AB} is solely dependent on the level differences and is in this case constant (Eqs. (3), (4), and (7)): applying the values $a_1 - a_2 = 2$, $b_1 - b_2 = 2$, and $n = 24$ to formulae 3 or 6 results in $E[MS_{AB}] = 96$. In contrast, when there is an interaction the summed treat-

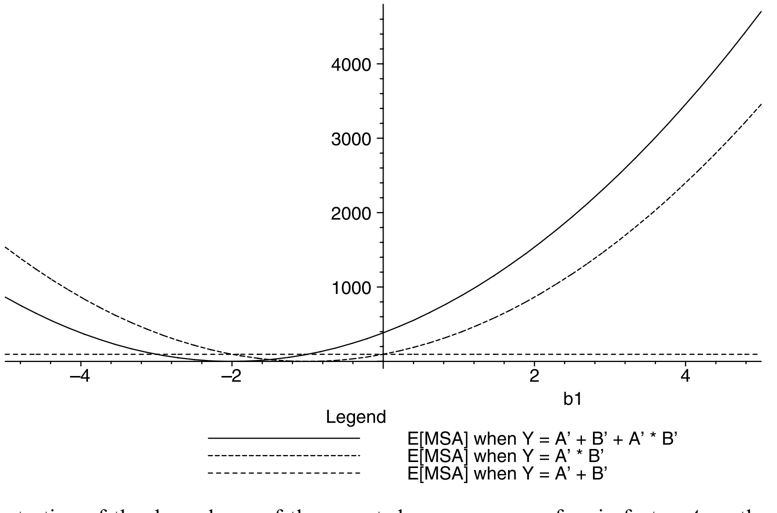

Fig. 3. Illustration of the dependency of the expected mean squares of main factor A on the treatment values b_1 and b_2. The figure shows, respectively, formulae 1, 4, and 8, while the treatment differences $a_2 - a_1$ and $b_2 - b_1$ are kept constant (2). The horizontal axis shows b_1 ranging from -5 to 5, the vertical axis shows $E[MS_A]$.

ment values of factor B have a strong influence on $E[MS_A]$. According to Eq. (7) $E[MS_A]$ is constant at 96 when there is no interaction ($Y = A + B$) and does not depend on the scale positions of b_1 and b_2.

Fig. 3 shows clearly that main effects depend strongly on the treatment values of the other variable when an interaction is involved. The relationship between $E[MS_A]$ and the scale values of variable B is quadratic. When $Y = A' * B'$, the main effect of A is zero only if the treatment values b_1 and b_2 are centered (in this case when $b_1 = -1$ and $b_2 = 1$). We notate the two consecutive values of b_1 and b_2 between round brackets. In this example, when $(b_1, b_2) > (0, 2)$ or $(b_1, b_2) < (-2, -4)$ the main effect strongly dominates the interaction contribution, and the power of the experiment to detect the interaction effect is considerably less than the power to detect the main effects. When $Y = A' + B' + A' * B'$, $E[MS_{\mathrm{A}}] = 0$ when $(b_1, b_2) = (-2, 0)$, while $E[MS_A] < 96$ in the area $(-3, -1) < (b_1, b_2) < (-1, 1)$ and $E[MS_A] > 96$ when $(b_1, b_2) < (3, -1)$ or $(b_1, b_2) > (-1, 1)$. Although the multiplicative term nearly always has the largest influence on the cell means to be analyzed, the MS_{AB} term remains constant in the different analyses. Therefore, the main effects often dominate the analysis results.

In many cases, the F value of the interaction component will be small compared with the main effects—for example in the *Ability* × *Motivation* relationship—when the treatment values are all positive. In that case, the interaction is called 'ordinal' in the literature. F_{AB} turns out to be relatively higher only, if signs are opposing. This type of interaction is called 'disordinal.' Comparison of the mean squares of the main effects and interaction effect leads to the conclusion that when $Y = A' * B' + e$,

$$E[MS_A] > E[MS_{AB}]$$

if

$$[n\{(b_1 + b_2)(a_1 - a_2)\}^2/4] + \sigma^2 > [n\{(a_1 - a_2)(b_1 - b_2)\}^2/4] + \sigma^2$$

or

$$(b_1 + b_2)^2 > (b_1 - b_2)^2,$$

and

$$b_1 * b_2 > -b_1 * b_2.$$

The main effect of A' will more rapidly be significant than the interaction effect if the values of B' have the same sign (both positive, or both negative). A similar conclusion holds true for B', but now

$$E[MS_B] > E[MS_{AB}]$$

leads to

$$a_1 \times a_2 > -a_1 \times a_2$$

so that the treatment levels of A' are decisive for the relationship.

5. Conclusions

In an additive model, the effects are solely dependent on the difference of the treatment values and their interpretation is straightforward. When the relationship is multiplicative in nature, the main effects intrinsically depend on both factors, as each of the main effects not only depends on the level difference of the corresponding factor, but also on the sum of the level values of the other factor. This means that whenever an interaction has been detected, the interpretation of the main effects is not simple nor straightforward, except in the case when both level values add up to zero. It also means that an interaction may not fully or solely represent the combinatorial effect of both independent variables. The dependency of the main effects on the treatment values of the other factor disappears only when both realized levels of each factor are centered around zero.

Because a multiplicative model can result in main effects as well as an interaction effect, and because the interaction is calculated as the effect that remains after both main effects have been partialled out, the interaction effect can be small compared with the main effects. This being the case, a test of the significance of the interaction effect may not be sufficient for the acceptance of a solely additive model. When an interaction is significant, there is sufficient ground for rejecting the additive model, but the opposite is questionable, as the experiment may lack the power to reject the H_0 of the interaction. As a consequence, the fixed ANOVA model is not a valid method to differentiate between additive and multiplicative effects.

6. Discussion

Commonly, a researcher uses a fixed design to study group differences as the result of manipulations. The independent variables are not observed and the research questions concern the group differences on the dependent variable. Our basic assumption is that the true values of a manipulated variable can be positive in nature, and that they are non-centered in the empirical world. Furthermore, we assume that these realized true factor levels are responsible for the effect on the dependent variable. For instance, we have assumed that even not manipulated control groups have some positive score on ability and motivation and that the manipulated groups have obtained a higher level of motivation and ability. Others have made the same assumption. For instance, Sackett et al. (1998) pictured Vroom's Ability × Motivation interaction with both ability and motivation ranging from 0 to high.

In a fixed ANOVA the values to represent the factor levels are arbitrary and represent the treatment strength in an indirect manner. When realized factor levels are non-centered and when the outcome is the result of a multiplicative contribution of both factors, main effects are to be expected that are relatively large compared with the interaction effect. These main effects are difficult to interpret as they represent more than the unique effect of the corresponding variable.

We have shown that when there are positive true factor levels but no combinatory effects, ANOVA does an excellent job in dividing the total variance of the dependent variable attributable to both factors and the error component. The problem arises when both factors have a multiplicative effect on the dependent variable. We have shown that the main effects in a multiplicative model $Y = A' * B'$ are intrinsically dependent on the sum of the level values of the other variable. Our main conclusion is that analysis of a fixed design is not suitable for studying interaction effects and cannot be used to differentiate between additive and multiplicative effects.

In the hypothetical situation where a researcher can establish true factor levels that add up to zero, main effects would be simple to interpret. In that case the influence of the other variable disappears, because the main effects would become zero when Y is solely dependent on the interaction $A' * B'$. Regrettably, it is not possible to establish centered treatment values a posteriori. The use of orthogonal coding of the independent variables does not guarantee that the true values of the factors are orthogonal, because these independent variables represent the true but unobserved factor levels in an indirect manner.

The phenomenon can be seen as a power problem in the detection of interaction effects. Of course, a power problem can be remedied by increasing the experimental power, for example by increasing the number of subjects, applying stronger treatments, reducing the amount of error through stronger experimental control, or more reliable measurement instruments. But although increasing power may make it possible to detect interactions, it does not make the main effects easier to interpret as they can be the possible result of non-orthogonal treatment values.

Clearly, the study of interaction can show us when effects of treatments are dependent on each other. Testing for interaction with ANOVA or multiple regression analysis allows us to falsify additive models to some extent, which is useful in itself. But in our view, it does not make much sense to interpret main effects and interaction separately. As Pedhazur (1997) said: "It would be well to remember that one administers treatments—not what is left after adjusting for treatment effects, which is what the interaction term represents." (p. 474). When treatments are not strictly centered, it is very difficult to interpret the way in which a 2×2 fixed effect analysis of variance allocates the effects over the three independent parts, the two main effects and their interaction. Rosnow and Rosenthal (1995) are justified in warning against disregarding main effects whenever an interaction has been detected. Our demonstrations have shown that the main effects indicated by ANOVA results can be the result of a combinatorial effect and should not be disregarded. We must add that in a fixed 2×2 design, small and even insignificant interaction effects should be taken seriously, and that one should not adapt the additive model too soon. Whenever an interaction is present, an interpretation directed at the total effect of both treatments may be the best course, as both main effects and interaction can result from the combination of both factors. We must also warn against accepting additivity simply because of lack of significance of the product term.

In real experimentation, the treatment values a_1, a_2, b_1, and b_2 are often not known, and often the variables are considered to be nominal. Wahlsten (1990), for instance, focuses on research into hereditary, where two different strains and two different environments are compared. The specific difference between the two strains that causes a given effect may be unknown, and even if it is known it may be difficult to measure the true difference between the two strains. The same reasoning can be applied to the differences between the two environments. As long as the true nature of the qualitative differences between the factor levels cannot be established, the interpretation of the interaction component as such will lead to ambiguity.

Knowledge of the nature and strength of the independent variables are a logical prerequisite for the interpretation of statistical analysis results. In general, it would seem advisable to assess factor levels, either by assessing the strengths of the treatments or by observing realized factor levels, and to use regression analysis for the study of multiplicative models. If the independent variables can be quantified to some extent, multiple regression analysis is indisputably the preferable analysis approach. Using measured treatment levels offers a better and more direct estimator of the resulting effect on the independent variable than arbitrary nominal values. Apart from the multicolinearity problem, extensively discussed in Bohrstedt and Goldberger (1969), Belsley (1991), and Pedhazur (1997), multiple regression analysis provides more accurate information of the contribution of the three different terms than can a fixed ANOVA design.

In discussions with colleagues, we have been offered several other suggestions for possible solutions. First, it is advised to fit the correct model, and if an interaction model is assumed, the interaction should be fitted first. Using SPSS, we have to discuss several approaches. Classic SPSS ANOVA, only available by the use of syntax, does not allow testing of the interaction effect separately. GLM does something quite different from SPSS ANOVA, as it attributes all degrees of freedom to the interaction effect when a single $A * B$ model is fitted. It tests whether there is any effect at all, but it is not very helpful in differentiating between an additive and a multiplicative model. Multiple regression analysis allows us to enter the separate effects in an ordered fashion, but the interaction effect remains relatively small, independent of the ordering of the variables. The latter is not surprising, as the smaller interaction is caused by the way it is calculated.

Second, we have been advised to use the methods of Searle (1987). Searle introduced the $R()$ notation, which is very useful in unbalanced designs. It offers no specific help for the discrimination of multiplicative and additive effects, as it leads to the same Sums of Squares in balanced designs such as our demonstrations, and the ambiguity between main effects and interaction effects remains.

Third, we were also often advised to center the independent variables. This advice is not very helpful, as the independent variables are centered (the values -1 and 1 are assigned), both in our examples and in the section about the calculation of the Sums of Squares. For abolishing hard-to-interpret effects, centered treatment strengths should have been obtained, but this is almost impossible to achieve in practice. Simple orthogonal coding of the independent variables is not sufficient. Centering is presented as the solution for various problems, including the issue of multicollinearity (Dunlap and Kemery, 1987). In contrast, Belsley (1991) has argued against the use of centering and Pedhazur (1997) warned that centering may obscure collinearity without reducing it.

The best advice is perhaps to improve on the knowledge of the nature and strength of each of the treatments, and to use measured independent variables whenever possible. Awareness of both the interpretational problems that are inherent to the fixed ANOVA model when an interaction is expected and the power problems concerning the rejection of the multiplicative model, should prevent jumping to conclusions as far as interaction effects are concerned. Additionally, we advise to focus interpretation on the total effect of both treatments whenever a multiplicative model is valid and not solely on the interaction effect. Further research on the true nature of both ANOVA and multiple regression analysis seems a necessity to us, despite the general usage of these techniques.

In this paper, we have assumed Vroom's model to be true and used it to demonstrate that ANOVA results may not indicate true interactions. It should be clear that we did not perform research on this subject ourselves. We have no strong opinion on the true relation among ability, motivation, and performance, but only warn of the possible misinterpretation of experimental results.

Appendix A. Demonstration $Y = A \times B$

TITLE reconstruction of performance $(Y) = \text{motivation}\ (A) \times \text{ability}\ (B)$ in Tables 1 and 2.

```
* open a file with 96 cases
GET FILE = '2^2 × 96.sav'.
* create orthogonal vectors a, b.
if (MOD($22CASENUM,4) = 0) a = −1.
if (MOD($22CASENUM,4) = 0) b = −1.
if (MOD($22CASENUM,4) = 1) a = −1.
if (MOD($22CASENUM,4) = 1) b = +1.
if (MOD($22CASENUM,4) = 2) a = +1.
if (MOD($22CASENUM,4) = 2) b = −1.
if (MOD($22CASENUM,4) = 3) a = +1.
if (MOD($22CASENUM,4) = 3) b = +1.
* create result according to the product relation, with a1 = 2, a2 = 4, b1 = 2, and
b2 = 4.
* and add independent measurement error the component.
compute err = normal(7).
Compute Y = (a + 3)*(b + 3) + err.
GLM Y BY a b
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/PRINT = DESC ETASQ OPOWER
/CRITERIA = ALPHA(.05)
/DESIGN = a b a*b.
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R COLLIN TOL CHANGE
```

/CRITERIA = PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT y
/METHOD = ENTER a/ENTER b/ENTER ab.
ANOVA Y BY $a(-1,1)$ $b(-1,1)$.

References

Belsley, D.A., 1991. Conditioning diagnostics: collinearity and weak data in regression. Wiley, New York.

Bohrstedt, G.W., Goldberger, A.S., 1969. On the exact covariance of products of random variables. Journal of the American Statistical Association 64, 1439–1442.

Cohen, J., 1968. Multiple regression as a general data-analytic system. Psychological Bulletin 70, 426–443.

Cohen, J., 1978. Partialled products are interactions; partialled powers are curve components. Psychological Bulletin 85, 858–866.

Cohen, J., Cohen, P., 1983. Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences, second ed. Erlbaum, Hillsdale, NJ.

Dunlap, W.P., Kemery, E.R., 1987. Failure to detect moderating effects: is multicollinearity the problem? Psychological Bulletin 102, 418–420.

Fisher, R.A., 1935. The Design of Experiments. Oliver and Boyd, Edinburgh.

Hayduk, L.A., Wonnacut, T.H., 1980. "Effect equations" or "effect coefficients": a note on the visual and verbal presentation of multiple regression interactions. Canadian Journal of Sociology 54, 399–404.

Hays, W.L., 1973. Statistics for the Social Sciences, second ed. Holt, Rinehart and Winston, New York.

Jaccard, J., 1998. Interaction Effects in Factorial Analysis of Variance. Sage Publications, Thousand Oaks.

Jaccard, J., Guilamo-Ramos, V., 2002. Analysis of variance frameworks in clinical child and adolescent psychology: issues and recommendations. Journal of Clinical Child and Adolescent Psychology 31, 130–146.

Jaccard, J., Wan, C.K., 1995. Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: multiple indicator and structural equation approaches. Psychological Bulletin 117, 348–357.

McClelland, G.H., Judd, C.M., 1993. Statistical difficulties of detecting interactions and moderator effects. Psychological Bulletin 114, 376–390.

Pedhazur, E.J., 1997. Multiple Regression in Behavioral Research, third ed. Harcourt Brace, Fort Worth.

Rosnow, R.L., Rosenthal, R., 1995. Some things you learn aren't so: Cohen's Paradox, Asch's Paradigm, and the interpretation of interaction. Psychological Science 6, 3–9.

Sackett, P.R., Gruys, M.L., Ellinson, J.E., 1998. Ability–personality interactions when predicting job performance. Journal of Applied Psychology 83, 545–556.

Saunders, D.R., 1956. Moderator variables in prediction. Educational and Psychological Measurement 16, 209–222.

Searle, S.H., 1987. Linear Models for Unbalanced Data. Wiley, New York.

Vroom, V.H., 1964. Work and Motivation. Wiley, New York.

Wahlsten, D., 1990. Insensitivity of the analysis of variance to heredity-environment interaction. Behavioral and Brain Sciences 13, 109–161.

Wahlsten, D., 1991. Sample size to detect a planned contrast and a one degree-of-freedom interaction effect. Psychological Bulletin 110, 587–595.