INTER-EMOTION COMPARISONS OF FACIALLY EXPRESSED EMOTION INTENSITIES: DYNAMIC RANGES AND GENERAL-PURPOSE RULES.

Armando Oliveira[1], Marta Teixeira[1], Isabel B. Fonseca[2], Eduardo Santos[1], Miguel Oliveira[1]
[1] *Institute of Cognitive Psychology, University of Coimbra, Portugal*
[2]*Faculty of Psychology and Educational Sciences, Lisbon, Portugal*
l.dinis@fpce.uc.pt

Abstract

The possibility of comparing qualitatively different dimensions on a common unit scale, either regarding its importance or its magnitude levels, is offered in functional measurement by the averaging model. Using distinct emotions as factors and different intensities of facially expressed emotion as levels in each factor, a general equal-weight averaging rule was found for all combinations of emotions in tasks that required subjects to judge overall intensity. Functional measurement of the factorial levels and factor weights followed, making possible to compare subjective dynamic ranges and spacing between levels of intensity across emotions. Several features of the data suggest the need to refine on the type of experimental task used, and a probable default character of the rule found.

One reason to want to compare the intensities of different emotions is practical as well as methodological. Whenever the issue of the differential effects of a particular type of emotion (e.g. fear *vs*. joy) is addressed, one should be capable of telling apart the outcomes of intensity from those of emotion quality, under risk of confounding. This is all the more so when the effects considered are likely dependent on intensity levels, as often is the case in practical matters. Pleading for the maladaptive character of fear, or distinguishing between the social consequences of joy and anger, to give a few examples, clearly requires that intensity aspects be accounted for. If no partial out of intensity differences is made, conclusions regarding emotion specificity may be essentially flawed. This point was first brought about in a systematic way by N. Frijda, who also made attempts at inter-emotion comparisons of intensity (Frijda, 1992). As an outcome of this, he issued cautionary remarks regarding available data and stressed the need for new methods to make such comparisons meaningful.

Inasmuch as distinct emotions are actually instances of qualitatively different variables, a principled solution exists in the framework of functional measurement, provided by the averaging model. A main feature of averaging theory is the possibility of separately estimating weight and scale value parameters (Anderson, 1981, 1982, 1996). Under certain conditions, which warrant for parameter's identifiability, it permits legitimate comparisons of both value and importance among the stimuli-variables embedded in an integration task. Typically, value estimates will be on a linear scale, with common unknown zero, meaning that differences between levels (spacing), although not absolute values, can be soundly compared across factors (Anderson, 1981).

A previous study employing this rationale was done with emotion lexicons (Oliveira et al., 2003). Emotion words from the same family (e.g., "joy-words") but expressing distinct intensities were used as factor levels, and subjects made to judge the combined intensity of word pairs pertaining to two different families (actually, averaging was prescribed to subjects). Near-additive integration patterns were found, allowing for the functional measurement of words in both emotion-factors. However, functional measurement could only be done at the single-subject level, since emotion-words were valued differently – and even disordinally - by different subjects. The range of stimulation granted by those words was also somewhat uncertain, given the problems involved in the sampling of emotion lexicons.

The present study follows the steps of that earlier one, with three main differences. (1) No combining operation is prescribed. (2) It includes subdesigns that allow for a proper estimation of functional weights and values in case the averaging rule is established. (2) It uses as stimuli emotion-conveying faces, which produce full agreement in rankings by intensity (within each emotion category) and offer naturally grounded dynamic ranges of stimulation. This last trait can be of help in ascertaining the importance of each emotion in the integration task, and also as a means to compare distinct emotions on the range of expression they are afforded in facial displays.

Method

Subjects

50 graduate students at the University of Coimbra, randomly allocated by groups of 10 to each of five tasks, took part in the experiment. They were all naïve regarding the topic under study and presented normal or corrected to normal vision.

Stimuli

Stimuli consisted of photos of faces selected from the databases *JACFEE and JACNeuF* (Matsumoto & Ekman, 1988). Faces were picked up by pairs, comprising a neutral and a very high intensity expression of a given emotion by the same individual. These two photos were then taken as endpoints for a digital morphing at equal steps of 1/3. The first morph obtained, just after the neutral stance, was taken as a "low" level of emotion expression, the following one as a "medium" level, and the most intense as a "high" level of expression. The choice of the morphing rate was dictated both by the need to ensure clear discrimination between levels and by evaluations of the "naturalness" of the expressions as representatives of low, medium and high intensity facial displays. Four emotions from Ekman's taxonomy (Ekman, 1993) were considered: fear, sadness, joy, anger.

Design and Procedure

The experiment involved 5 integration tasks, all of them obeying a 3 x 3 repeated measures complete factorial design (n = 10). Each task used as factors two distinct emotions and as levels three intensity levels of emotion expression: low, medium, high. The two corresponding one-way subdesigns were always included, as a means to decide between averaging and adding operations if an additive pattern came about, and also to obtain uniqueness of weight and scale parameters in case averaging prevails (see Anderson, 1982). Since four emotions were considered, 6 tasks would be needed to have them combined two by two. However, an accidental loss of files prevented data of the sixth experience (*Anger* x *Fear*) to be analysed and reported here.

Each stimulus was presented twice (with reversed placement of the two expressions in each pair) on a computer display, for a total of 30 randomized experimental trials. Subjects were run individually, after a variable number of training trials, and asked to judge overall intensity conveyed by pairs of expressions on a 0-20 rating scale. Instructions were read aloud and open to questioning during the entire training period.

Results

Integration algebra

As illustrated by Fig. 1, all factorial plots exhibit near-parallelism, the signature of a general additive rule (by virtue of the parallelism theorem of IIT, which simultaneously supports the

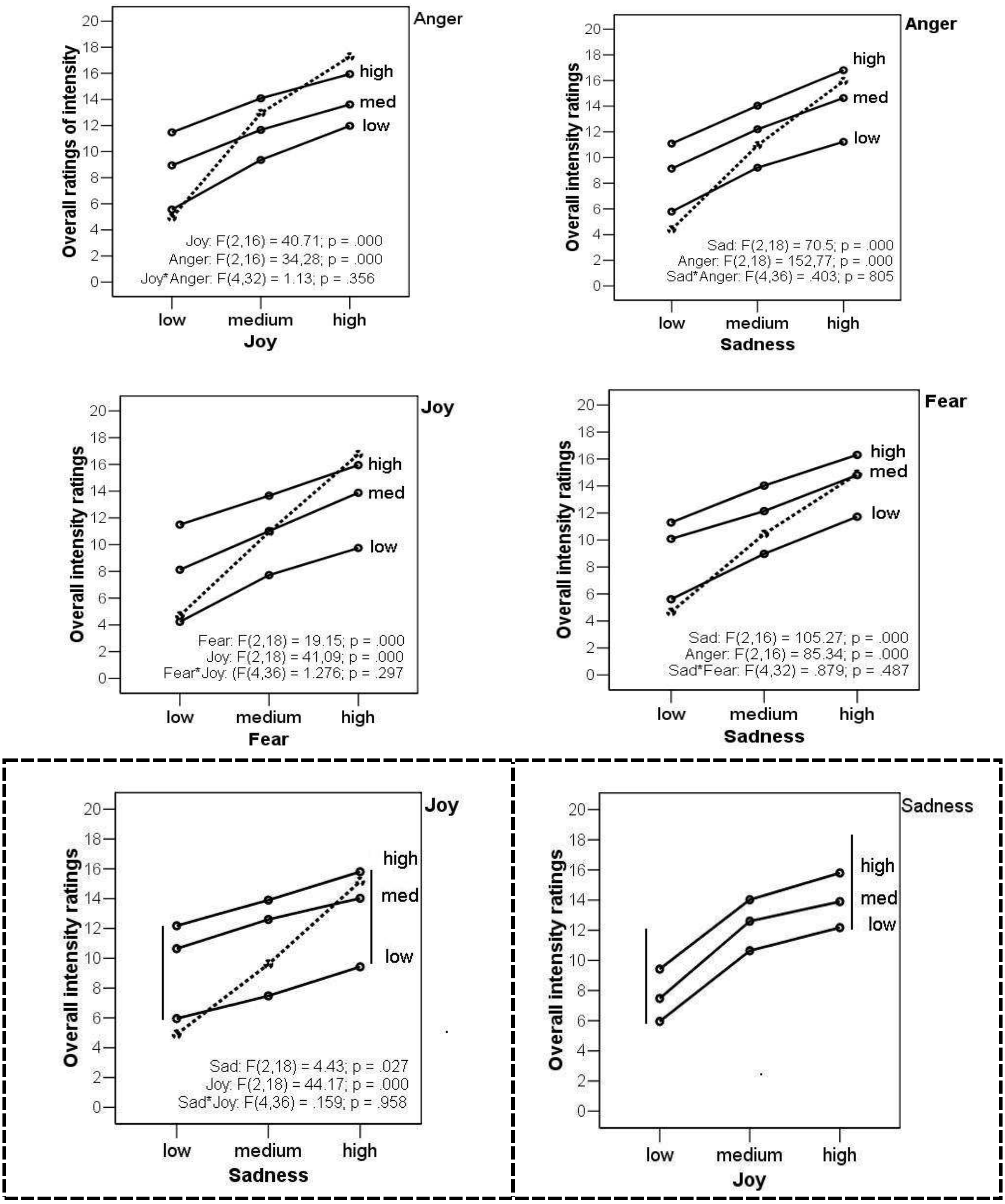

Figure 1. Integration patterns of intensity levels of distinct emotion categories (two by two).

response scale as linear) (Anderson, 1981, 58-59). The dashed line added to the graphics represents judged intensity of levels of the emotion in abscissa presented in isolation. The clear cross-over obtained in all cases rules out addition and establishes averaging as the proper model in the present case (Anderson, 1982, 215-217).

Statistical analysis concurs with visual inspection of the patterns. Significant main effects were always obtained, while the interaction term never reached significance (lower p value = .297; see values in Fig 1). This is what could be expected from a general additive model, which implies null interaction. Taken all together, data lend consistent support to an equal-weighting averaging operation (i.e., same weight across levels within each factor).

Functional measurement

The finding of an averaging rule with equal weighting, given that averaging then reduces to an additive-type model, allows using marginal means of responses as functional values of the stimuli. This would be enough for purposes of quantifying expressions within each factor on an interval scale. However, it isn't enough for interval comparisons across factors. The point is that marginal means correspond only to gross stimuli values, reflecting the mixed effects of weight (constant within each factor, but eventually differing between factors) and value parameters. Direct comparisons of spacing between levels across factors would thus only be legitimate if both factors shared the same weight. In general, between-factors comparisons require weights and "proper" scale values to be separately estimated through suitable procedures. This was achieved via AVERAGE, a program specifically developed for analysis of the averaging model (Zalinski & Anderson, 1987)

The two graphics in the bottom panels of Fig. 1 (surrounded by dashed borders) illustrate the problem. They plot the same data, corresponding to the *Sad* x *Joy* experiment (*Joy* as curve parameter on the left panel and *Sad* on the right one). Vertical spread is clearly superior for Joy than for Sad (see the vertical bars of equal length added to the graphics). However, no unambiguous interpretation can be made of this. If one accepts that selected stimulus cover the relevant dynamical range of both emotions, the difference in spread might still mean different things: that the expression range of Joy is greater, or that Joy had greater weight/importance than Sad in the task (with equal or even smaller range). Unequivocal interpretation of these data can only be arrived at by telling apart the respective roles of weights and scale values.

Parameter estimation

The AVERAGE program was run with data from each of the five experiences (with one-way subdesigns as a requisite for uniqueness of estimation). The equal-weighting model was selected as an option and starting scale values inputted were symmetrically bounded at 20% of the full response range (Zalinski & Anderson, 1991). Equal starting weights were given to each factor and bounded according to the general guidelines provided in the AVERAGE manual (Zalinki & Anderson, 1987). Estimation was done separately for each subject. This allows to get round bias found in weight estimates from group means (Anderson, 1982, p. 94) but, most of all, affords a distribution of parameter values which can be used for statistical tests of differences. The fit of the equal-weighting averaging model was always assessed by running ANOVAs (with the same 3 x3 main design) on the residuals left by the model predictions. In every case, no significant main effects or interactions were found, which indicates that the model accounted for all systematic sources of variance.

Two distribution of weights derived this way were thus obtained in each task (one for each factor-emotion involved). Mean values of the distribution were as a rule very similar,

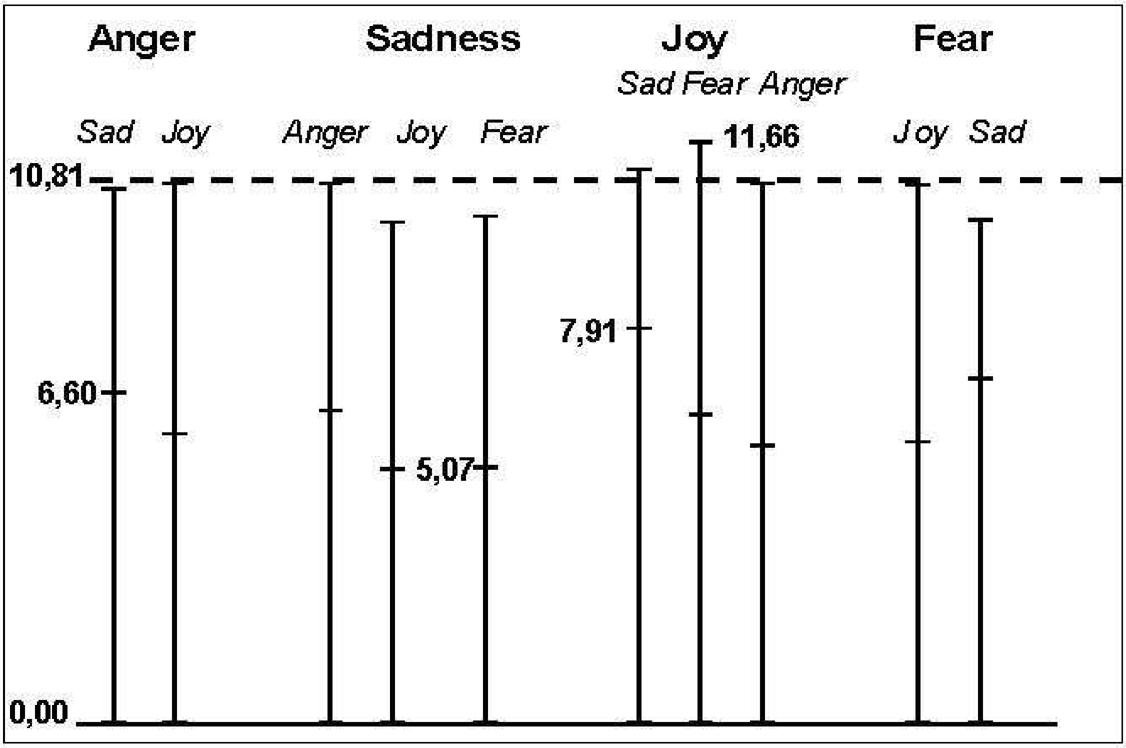

Fig. 2. Functional scales of perceived emotion intensity obtained from value estimations. (italics identify the other emotion in the integration task from which the scale was derived)

exception made to the *Sad-Joy* task (5.97 vs. 8.19 [arbitrary units]). Paired-sample *T* tests failed to reveal significant differences between these weight distribution. Expressed emotions thus seemed to generally share the same weight/importance in each task. As for scale values, the same general procedure was followed. Distributions of total functional ranges were compared between emotions within each task, as well as distributions regarding the medium level of expressions. In all cases, the lowest value in the scale (lowest intensity of emotion expression) provided the ground anchor of the range, being equated to a conventional zero.

Figure 2 graphically represents the functional scales obtained for each emotion from different tasks (italics allow to identify the particular task from which a scale was derived). Suggestions of higher ranges of dynamical expression for Joy regarding Sadness and Fear, as well as of higher "low-medium" differences in Joy regarding Sadness, are the most apparent trends. This was only partially supported by statistical analysis. Distributions of "low-medium" spacing were significantly different for Joy and Sad (t (9) = 3.318; p = .009), while differences in entire functional range between Joy and Fear (J-F) could only qualify as marginally significant (p = .064). No other significant differences were found.

Conclusions

A general equal-weighting averaging rule is at work in the joint appreciation of intensities of distinct emotions conveyed by faces. Minor deviations from parallelism, such as the slight convergence to the right in the Joy x Anger pattern (top-left panel on Fig. 1) do not seem of consequence; not only are they not supported by statistical analysis (under the reservation of available power) as the fit of the equal-weighting model on the basis of residuals left has proven good in all circumstances.

This result sets the requisite conditions for functional measurement of expressed intensities within each emotion and, as a prime goal, of sensible comparisons of intensity between emotions. Outcomes of functional measurement obtained from these tasks tended to reveal equal importance of emotions conjoined in each task (with a possible exception for *Sad* vs *Joy*: joy benefiting from greater weight), as well as relatively uniform full functional ranges across emotions (again, with a possible exception for Joy, which showed a tendency for increased ranges).

These homogeneous trends may be an upshot of the very nature of the task, which, unlike natural settings, promotes equal consideration of both expressions. Also, given that averaging affords a general-purpose device for integration of multiple sources of information, it could be that the model found actually corresponds to a default rule. This is not an argument against the employed rationale as, for measurement purposes, even prescribed rules can be rightfully used; still, it stresses the convenience of moving toward tasks where blends of emotions and intensities can be less artificially and more substantively evaluated. A possible path would be to transfer this same approach to the level of facial elements such as the Action Units distinguished in the FACS system (Ekman & Friesen, 1978), thus allowing judgment to operate within the more natural context provided by a single facial expression.

The small number of subjects involved in each task qualifies these results as provisional. However, great inter-individual consistency was revealed by single-subject analysis. Also, the fact that each task was done by a different group of subjects (for practical reasons) can be a limiting factor to be overcome in subsequent work.

References

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic Press.

Anderson, N. H. (1982). *Methods of information integration theory.* New York: Academic Press.

Anderson, N. H. (1996). *A functional theory of cognition.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Ekman, P. (1993). Facial Expressions and Emotion. *American Psychologist*, 48, Nº 4, 384-392.

Ekman, P. & Friesen, W. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.

Matsumoto, D. & Ekman, P. (1988).*Japonese and Caucasion facial expressions of emotion (JACFEE)*. San Francisco, CA: San Francisco State University, Intercultural and Emotion Research Laboratory.

Frijda, N., Ortony, A., Sonnemans, J. & Clore, G. L. (1992). The complexity of intensity. Issues concerning the structure of emotion intensity. In M. S. Clark (Ed.), *Review of Personality and social psychology* (Vol. 13, pp. 60-89). Newbury Park, CA: Sage.

Zalinsky, J. & Anderson, N.H.(1987). *AVERAGE Program & Manual*. San Diego: University of California.

Zalinski, J. & Anderson, N. H. (1991). Parameter estimation for averaging theory. In N. H. Anderson (Ed.), *Contributions to information integration theory. Vol. I: Cognition* (pp. 353-394). Hillsdale, NJ: Lawrence Erlbaum Associates.

Acknowledgment

This research was partially supported by grant POCTI/PSI/41235/2001 of the Portuguese Foundation for Science and Technology