JOINT MODEL-PARAMETER VALIDATION OF SELF-ESTIMATES OF VALENCE AND AROUSAL: PROBING A DIFFERENTIAL-WEIGHTING MODEL OF AFFECTIVE INTENSITY.

Armando M. Oliveira[1], Marta P. Teixeira[1], Isabel B. Fonseca[2], 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

Based on a hypothesised differential weighting model of integration of valence and arousal into overall affect intensity, self-estimates of valence and arousal where inputted to the model and derived predictions of intensity compared to observed judged intensity. Given that separable factorial manipulation of valence and arousal was hampered by the stimuli used (affective pictures from IAPS), self-estimation was mandatory and no functional parameters could be independently established as independent criteria. Two estimation methods were employed: absolute weight/value estimation and part-worth estimation. Moderate adjustment of the model predictions offered partial support to both the assumed averaging rule and the self-estimates employed. As for self-estimation methods, part-worth and direct separate judgments of value and importance seem to have behaved at an equivalent level.

Valence and arousal have been recurrently identified as major organizing dimensions of affective space (Larsen & Diener, 1992). Its relations to the intensity of affect/emotion have also been a topic of debate. The idea that emotion intensity may act as an overarching dimension to which a host of other dimensions contribute as determinants has been put forward, under different forms, by a number of authors. As one of them, R. Reisenzein specifically suggested that affective intensity be equated with the joint total contribution of valence and arousal to a feeling or emotion (Reisenzein, 1994). Even if this suggests a rule for the integration of valence and arousal into overall intensity, the meaning of "total" was left unspecified in Reisenzein's proposal (e.g., the outcome could be the one of an addition, a multiplication, averaging, etc.)

This study is intended as an exploratory inquiry on the existence of such a rule. It hypothesises a differential weighting averaging operation as the most plausible candidate, in case the rule actually exists. This assumption can be justified on several grounds: the averaging rule with differential weighing appears to be the most ubiquitous device for information integration (Anderson, 1982, 1996); empirically, prior studies addressing the integration of pictorial representations of valence and arousal, with judged global intensity as response, have found an averaging rule (Oliveira *et al.*, 2005); considerable evidence, behavioural as well as psychophysiological, points to extensive occurrence of differential weighing in affective/evaluative matters (e.g, the "negativity" bias": see Ito *et al.*, 1998).

To go beyond mere representations of valence and arousal poses a problem, since valence and arousal will typically became associated in each single stimulus, precluding its separate manipulation as factors in an experiment. As a consequence, subjects must also estimate factors' levels, and not just the final judged dimension, to bring to light the integration pattern. In such situations, referred to as "joint model-parameter validation", both model and estimated parameters come under test at the same time. To obtain a good model fit grants simultaneous support to the model and to the estimations used. In case no acceptable fit is obtained, not much will have been gained (Anderson, 1982, 279-282).

Two distinct estimation methods for valence and arousal were used in this study: part-worth estimation, which asks to partition the overall evaluation of a stimulus into the respective contributions of each factor, and direct judgments of weight and value. The second method enables to input required self-estimated parameters to the model, and thus to assess its fit. The former one, in case the model verifies, offers the possibility of comparing performance of the two distinct estimation methods.

Method

Subjects and stimuli

27 graduate students at the University of Coimbra took part in the experiment: 15 of them were assigned to the part-worth estimation task, and 12 others to the task involving direct judgments of value and importance.

Stimuli were pictures from the *International Affective Picture System* (Lang & Bradley, 1988), selected to obtain six distinct combinations of arousal and valence according to a 3 valence (negative, neutral, positive) x 2 arousal (low, high) design. 36 pictures were used in total, 6 for each combination. Given that each experimental condition is defined by the pooling of different stimuli, "univocal" definition of levels of the factors was formal and not substantively achieved.

Design and procedure

A general repeated measures 3 x 2 design, with no replications, was used in two separate sessions, both in the "part-worth" and the "judgement of value-importance" tasks. Pictures were randomly presented one by one on a computer screen located 50 cm ahead of the subject. A 0-20 rating scale was always used for response. A training period was ensured with pictures not included in the experiment.

(1) Judgment of importance and value: in the first session subjects were made to estimate valence, arousal, and intensity values for each picture ("in-context" estimation (Anderson, 1982)). The order of the estimations was handled by a latin square which produced six ordering sequences (even balance wasn't actually achieved because of inadequate number of subjects). In a second session, occurring the following day, subjects had to separately estimate the importance of valence and arousal to the global impact produced by each picture (again on 0-20 rating scales).

(2) Part-worth task: first session as in the previous task. In the second session subjects were faced with the intensity estimates they previously gave to each picture, and asked to partition that value into the respective contributions of valence and arousal. As a requisite condition, than, these contributions (part-worths) had to sum to the intensity estimated value.

Results

Direct judgments of value and importance

The leftward graphic on Fig. 2 displays judged intensity and intensity predictions from the differential-weighting averaging model, using self-estimated values and weights (of valence and arousal) as inputs. Positive valence conditions exhibit near to perfect agreement, irrespective of arousal levels. To a lesser degree, the "negative-high" and "neutral-low"

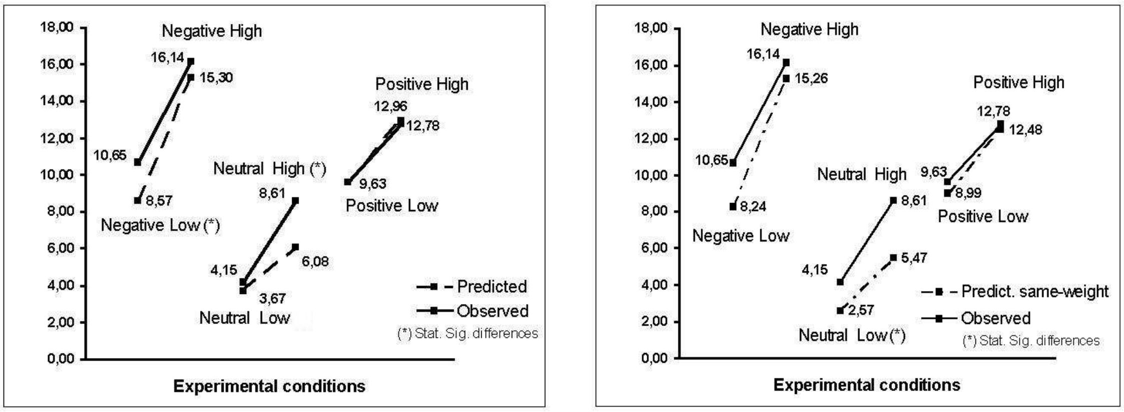


Fig 2. Predicted *vs.* judged overall intensity:
Left: using self-estimated weights *Right:* using equal (0.50) weights

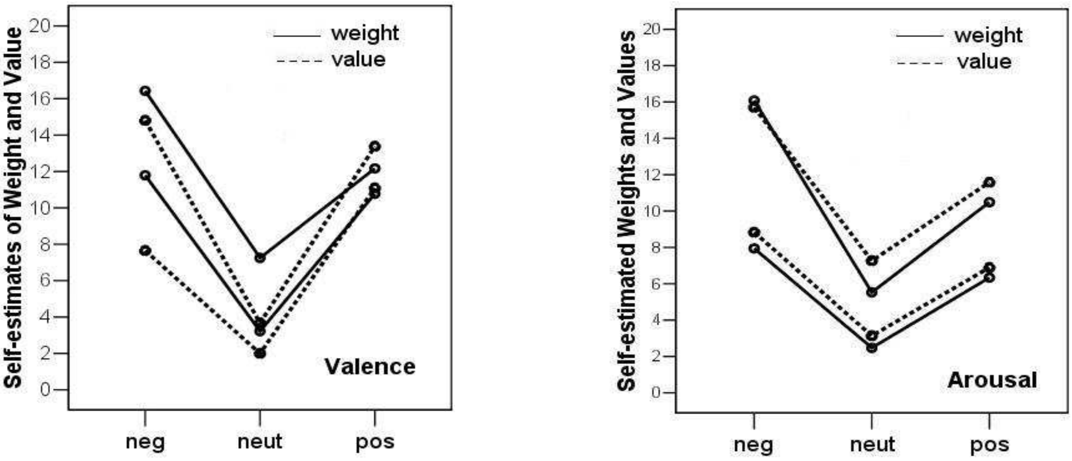


Fig. 3. Comparisons between weight and value self-estimates *Left*: valence. *Righ:* arousal

conditions similarly disclose reasonable fit. The most obvious discrepancy occurs at the "neutral-high " condition, closely followed by the "negative-low" one. The rightward graphic compares observed values to predictions obtained from a simpler additive-type model (equal weights across levels in each factor and across factors), which might as well be in Reisenzein's mind. On the whole, this alternative model shows worst fit than the previous one, most sensibly at the positive and neutral conditions.

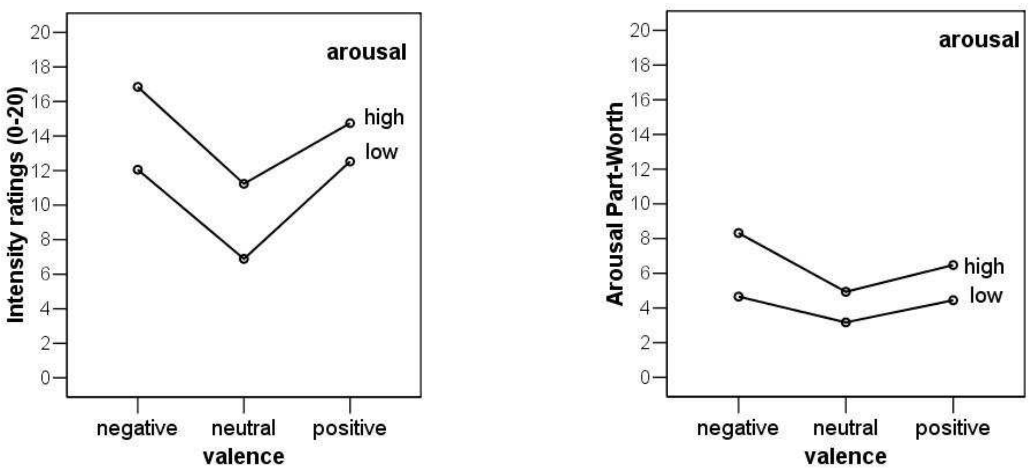
Given that predictions were derived at the individual level (just as estimates of overall intensity) statistical tests can be performed between judged and predicted intensities. Significant incongruities are signalled in the graphs, and concur with the outcomes of visual inspection. Comparative advantage of the differential weighing model is also corroborated by the statistical comparisons, which reach significance at the "neutral-low condition" in the additive model, but not in the differential-weighing model. Some degree of positive contribution of self-estimates of weight to the fit obtained thus appears warranted.

Figure 3 reveals very similar values of weight and value estimates for arousal and, to a lesser extent, for valence. This may be caused by difficulties in distinguishing value and importance, a problem that doesn't affect the part-worth method (where weight an scale value are actually merged in the total input of the part to the whole). It can moreover be seen that no true independence between levels of valence and arousal was obtained with the selected pools of stimuli. This plausibly reflects the fact that valence and arousal are not truly independent, in which case it also sets an internal limit for model fitting in terms of algebraic models.

Part-worth estimation of valence and arousal

Figure 1 presents the factorial plots corresponding to intensity judgments and equivalent plots for the part-worth responses. Differences between lines (vertical spread) are the important element to consider, since they entail a validity check of self-estimated part-worths. If the effect on intensity of the augmentation of arousal is well captured by self-estimated part-worths of arousal (its judged contribution, as a part, to the whole), than differences between lines should be the same in part-worth and intensity plots (see Zhu & Anderson, 1991). This applies likewise to valence.

Graphs show a varying extent of agreement as a function of the levels of the second factor, in the abscissa. This is not a problem for the validation rationale, which doesn't require constant differences between lines, and applies to each determining point in the curves. As the most apparent trends, larger divergences in arousal part-worths occur at the neutral level of valence, which is involved as well in the main discrepancies found for valence (more clearly so at the higher level of arousal).


Intensity: vertical spread of arousal

Part-Worth estimation (arousal)

Intensity: vertical spread of valence

Part-Worth estimation (valence)

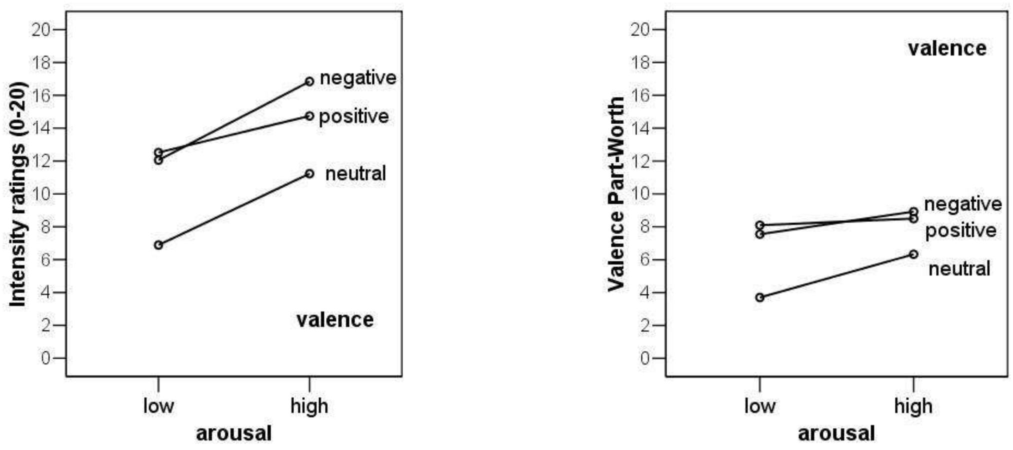


Figure 1. Factorial patterns for overall Intensity and Part-worth estimates:

Table 1. **Comparisons between self-estimated and intensity derived part-worths**

		Judged Part-Worths (1)	**Part-worths from Intensity (2)**	**(2) - (1)**	***Sig.* *t* (14)**
	Valence				
Low Arousal	Neg-Neut	3.85	5.16	1.31	n.s
	Pos-Neut	4.39	5.76	1.31	n.s
	Neg-Pos	1.63	2.38	.75	n.s.
High Arousal	Neg-Neut	3.24	5.61	2.37	.015
	Pos-Neut	3.03	3.53	.50	n.s.
	Neg-Pos	2.99	2.92	-.07	n.s
	Arousal				
Neg. Valence	Hig-Low	3.66	4.79	1.13	n.s
Pos. Valence	Hig-Low	2.02	2.22	.20	n.s
Neutral	Hig-Low	1.76	4.3	2.54	.000

Differences between lines for each determining point (abscissa) are numerically expressed in Table 1 (mean values). They confirm what was said for graphics, but with the analytical aid of statistical comparisons. Since differences between estimated and intensity derived part-worths were gathered for each subject, it is possible to statistically compare their distributions (paired *t* tests reported on the rightward column). Concerning valence part-worths, only the negative-neutral spacing in the high-arousal condition differed significantly from the criterion. Regarding arousal, part-worth spacing in the neutral condition was the only significant departure from part-worth given by intensity. On the whole, the part-worth method appears to perform at a very much comparable level to the judgment of weight and importance.

Conclusions

Results regarding the adjustment of predictions from the differential weighting averaging model to judged intensities are mixed. While they graphically show better overall agreement than simple equal weighting among all levels and factors (see Fig. 2), the statistical patterns of deviations do not differ too much between models. Almost perfect predictions were derived from the differential weighting model for positive valence conditions. By and large, outcomes are partly supportive of both the contribution of self-estimated weights to improved fit and of the probed, fitted model.

The similarity of weight and scale self-estimates on several conditions (Fig. 3) may be signalling some degree of confusion between value and importance, a danger incurred by self-estimation methods which require that they be explicitly distinguished (Anderson, 1982, Anderson & Zalinski, 1991). In that regard, arousal appears as more prone to the blending of weight and value than valence. Similar analysis made as a screening for possible confusions between intensity, valence, and arousal (considering both weight and value estimates), revealed rather distinct factorial patterns in all cases. Subjects thus seem to be able to make good sense of the demand to judge each dimension as distinct from the others.

Part-worth outcomes do not bear direct implications to the “model validation” issue, since they do not sort out values and weights to be used in predictions. However, once assumed the model is sound, they make possible to compare the relative efficacy of the two estimation methods employed (Zhu & Anderson, 1991; Anderson & Zalinski, 1991). In light of the evidence gathered, no clear-cut choice between the two methods appears justified.

The many downsides of this exploratory study stress the need to find ways of experimentally handling valence and arousal as separate factors. The probing of an integration rule and the estimation of model parameters still gets its gold standard from the validity criterion provided by an independently assessed integration pattern.

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. & Zalinski, J. (1991). Functional measurement approach to self-estimation in multiattribute evaluation. In N. H. Anderson (Ed.), *Contributions to information integration theory. Vol. I: Cognition* (pp. 144-185). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ito, T., Larsen, J., Smith, N., & Cacioppo, J. (1998). Negative information weights more heavily on the brain: The negativity bias in evaluative categorization. *Journal of Personality and Social Psychology, 75(4)*, 887-900.

Larsen, R. & Diener, E. (1992). *Promises and problems with the circumplex model of emotion*. In M. S. Clark (Ed.). *Emotion* (pp. 25-59) Newbury Park, CA: Sage Publications

Oliveira, A., Fonseca, I., Teixeira, M., & Simões, F. (2005). A functional measurement approach to the Self-Assessment Manikin. In J. S. Monahan, S. M. Sheffert, & J. T. Townsend (Eds.). *Fechner Day 2005. Proceedings of the Twenty-First Annual Meeting of the International Society for Psychophysics* (pp. 251-256). Mt. Pleasant, MI: The International Society for Psychophysics.

Reisenzein, R. (1994). Pleasure-Arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology, 73, 6,*1313-1329.

Zhu, S.-H. & Anderson, N.H. (1991). Self-estimation of weight parameter in Multiattribute analysis. *Organizational Behavior and Human Decision Processes,* 48, 36-54.

Acknowledgment

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