Are discrete emotions goal-derived or taxonomic categories? A study based on lexicon*

Armando M. Oliveira**
Eduardo R. Santos**
Marta Teixeira***
Isabel Barahona da Fonseca****

Abstract: Since the notion of typicality (graded structure) was first proposed, several of its possible determinants have been identified. Moreover, the distinction between common taxonomic categories and goal-derived categories has been paralleled with a suggested difference in their respective sources of typicality ("central tendency" and "ideal"). Building on this, previous studies using face materials have found evidence for a goal-derived nature of typicality in the domain of facial expressions of emotion. The work presented is partly aimed at assessing the generality of these conclusions by resorting to emotion-words (vs. faces). On the other part, it purports to highlight the relations of "ideals" and "central tendency" to "intensity" as a major dimension of both felt and expressed emotions. Outcomes found point to a close relationship of "ideal" to "central tendency", and to a variable relation between "ideal" and "intensity", depending on the specific emotion categories.

Key-words: typicality, ideal, intensity, emotion expression

Resumo: Na sequência da proposta da noção de tipicalidade (estrutura graduada), foram identificados alguns dos seus possíveis determinantes. Adicionalmente, diferentes determinantes de tipicalidade (e.g. "tendência central" e "ideal") foram associados a diferentes tipos de categorias (e.g., "taxonómicas comuns" e "derivadas-de-fins"). Com base neste conjunto de distinções, alguns estudos anteriores concluíram por uma tipicalidade assente em determinantes "ideais" no terreno das expressões faciais das emoções. O presente trabalho procura em parte avaliar a generalidade destas conclusões, recorrendo a um novo tipo de material (léxicos emocionais). Mais especificamente, porém, propõe-se examinar as relações entre "ideal", "tendência central" e "intensidade emocional". Os resultados encontrados sugerem uma relação estreita entre "ideal" e "tendência central", e uma relação variável entre "ideal" e "intensidade emocional", dependente das categorias emocionais consideradas.

Palavras-chave: tipicalidade, ideal, intensidade, expressão emocional

On indices and terminology

The first issue to be dealt with concerns terminology. We will be using the following variable names: "typicality", "central tendency", and "ideal". Each of these terms obeys a general definition, but they all allow for different meanings according to the concrete procedures in use to collect quantified indices. We will thus start by specifying the procedures we used and their supporting rationale, and by additionally comparing them to those employed by Hortsmann (2002) in a study with a similar scope concerning faces.

Typicality

In its most general sense, stemming from Rosh pioneering studies (1973; 1975), typicality refers to a graded structure underlying both category membership and category non-membership. For example, "robin" is a better (typical) member of the""birds" category than "ostrich"; on the other hand,""dog" is a better (typical) non-member of the category "fishes" than "whale".

At least two main types of questions can be used to elicit typicality judgments, which are not equivalent (Barsalou, 1985). The first one directly asks subjects about "how typical" an exemplar is; as Barsalou points out, this may lead them to emphasize "frequency of instantiation" (i.e, how often an exemplar occurs as an instantiation of the category). The second one asks subjects about "how good-an-example" an exemplar is; it is, according to Barsalou, freer from emphasis on specific determinants of typicality. It can be seen that the assessment of typicality suffers from ambiguities, and that these ambiguities depend on the kind of determinants over which typicality judgments are made to rely.

The subjects in our study had to rate "how suitable" each word-exemplar was to represent a given emotion. This is not so far from the "goodness of exemplar" instructions; also, as will become clear below, it closely addresses what Horstmann calls an "ideal" dimension (2002, 299).

Central Tendency (CT)

Family resemblance has been traditionally considered the major determinant of typicality (Rosch and Mervis, 1975). It can be envisaged from different angles, one of them being as all sorts of "central tendency" information in a kin-statistical sense (e.g., mode, median, mean, etc.). The cornerstone of this standpoint is the interchangeability between an exemplar's similarity to central tendency and its average similarity to all members in the category (Barsalou, 1983).

Horstmann embraced this operational understanding of family resemblance as "distance to CT" in his paper. Since he used in his study drawings of five facial expressions varying across five levels of intensity, CT scores were obtained by making participants judge the similarity of all pairs of faces within each category: the average similarity of each face to all the others in the category was then calculated and taken as an index of CT (Horstmann, 2002, 299-300). This is just one possible way of acknowledging the functional equivalence between average similarity and similarity to central tendency. It is also entirely dependent on explicit judgment and on the assumption of a one-dimensional continuum of similarity.

The CT indices we used were obtained differently. On the basis of the number of times each pair of emotion-words in a category has been jointly attributed to it by a number of subjects (n = 50), an input matrix of proximities was built up (thus,

one for each emotion category). This matrix, interpreted as a matrix of similarity between word pairs, was then treated through *multidimensional scaling* (ALSCAL) constrained to a single dimension solution. The normalized coordinates of that dimension, with zero as its central point, offer an alternative index of distance to CT or "family resemblance". The choice of constraining the MDS solution to one dimension introduces of course an arbitrary element. Mostly, it was decided on the following reasons: (1) to keep up with the one-dimensional assumption of usual measures of CT; (2) to offer some means of evaluating this assumption, basically trough checking the interpretability of the one-axis solution; (3) to offer a provisional unified framework across different emotions.

Ideal
Generally, it can be defined as any property, characteristic or dimension that an exemplar must possess in order to accomplish a goal associated with its category. As such, it may also offer a basis for graded structure (typicality). Barsalou exemplifies with «zero calories» as an ideal for the category «foods to eat on a diet» (1985, 630). In practice, there may be more than one ideal for each category; the typical question to be asked is how much (what amount) of the ideal is present in the exemplar.

The "ideal" scores in Horstmann study were obtained by asking subjects how suitable to express a given emotion a given facial expression was. The idea that lied behind was that "expressing emotion" constitutes the tacit goal served by facial expressions of emotion (Horstmann, 2002, 299). With minor differences in phrasing, it can be noticed the close resemblance

to the question we used to assess typicality, which means that we won't be handling in our study separate measures of "typicality" and "ideal", but just a single "typicality-as-ideal" index. In fact, differently from Horstmann, our main concern was not to establish the source of typicality in the realm of emotion expressions, but to examine the relations of both "typicality-as-ideal" and "central tendency" to *intensity* as a major dimension of either felt or expressed emotions.

Goal-derived and common taxonomic categories

The distinction between taxonomic and goal-derived categories has roots in the work of Barsalou (1983) on *ad hoc* categories. These later are constructed to achieve novel goals, and correspond therefore to labile structures not yet established in memory. They can nevertheless become engrained through use; as a result, goal-derived categories encompass both *ad hoc* and former *ad hoc* categories whose primary function is all the same to serve a goal.

Just like taxonomic categories, goal-derived ones exhibit graded structure, so the question arises whether determinants of typicality are different or the same for both types. Adding to a number of reasons (mainly of a functional nature) to expect them do be different, Barsalou has gathered evidence for a privilege of CT in taxonomic categories and of ideals in goal-derived ones (1983; 1985; 1987). These findings are however to be tempered with the following observations: (1) a graded structure may be simultaneously determined by more than one factor; (2) the determinants of a particular graded structure can change with context; (3)

rather than reflecting invariant structures, typicality seems to reflect people's ability to construct concepts (cf. Barsalou, 1985; 1989).

To sum up, although we may distinguish ideals and CT on ground of their favored association with different types of categories, there is no way of setting up a clear-cut boundary. The same applies to the very distinction between common taxonomic and goal-derived categories, which allows for no definite frontier. As orienting guidelines, Barsalou proposes the following differences: (1) common taxonomic categories are based on clusters of co-occurring properties and thus reflect the correlational structure of the environment, while goal-derived categories (such as "things to take on a journey") usually don't; (2) common taxonomic categories are often used to classify or represent kinds of entities, while goal-derived categories are normally used to achieve goals (such as in planning); (3) common taxonomic categories are highly familiar categories, with a biological or artifactual origin, well established into cultural knowledge, while goal-derived ones are not necessarily so.

The Horstmann's study (2002)

Horstmann's study embodies a different logic from those of Barsalou, who aimed at demonstrating that ideals can determine graded structure; with that in mind, he used specifically tailored goal-derived categories that he contrasted against well-known taxonomic categories. Horstmann, on his turn, deals with a single preexisting category (facial expressions of emotion) that, in light of the above-presented guidelines, might as well qualify as taxonomic or as goal-derived in nature. The

issue at stake was whether, on the basis of the observed determinants of typicality (namely "ideals" or CT), a decision could be reached as to the specific nature of the category (Horstmann, 2002, 298). The author concluded, on analysis, that ideal and not CT determined typicality, implying that categories of facial expressions are goal-derived in nature. This conclusion should of course be looked at carefully, considering the warnings about a flexible use of determinants and the potentially hybrid character of some categories,

One important issue in the Horstmann study concerns the role of *intensity* and its relation to the CT index. This index has been computed, following Barsalou, as an average of the similarity ratings obtained by each exemplar (after being paired with all the others in the category). Given the perceptive salience of intensity in the particular materials used – schematic drawings of facial expressions that varied in intensity – it is no surprise that CT corresponds to median intensity exemplars, while "ideal" varies linearly with intensity. The point to be made is that a less perceptually constrained similarity measure might allow for different results. We tried to arrive at such a measure (1) by resorting to different emotion-expressive materials (words) and (2) by adopting a more indirect way of computing the CT index, through MDS techniques.

Also, as Horstmann himself acknowledges, his stimuli materials are entirely made of pure emotional expressions, excluding blends. This is something that can only be achieved with faces, and would be highly implausible with words. A possible consequence is that, if we turn out to facial blends, or to materials such as lexicon, we might get increases in intensity accompanied by decrements in typicality. The empirical study that follows can be

partly conceived as an assessment of the generality of Horstmann findings. More centrally, however, its goal shifts away from the issue of typicality to specifically address the relations of *emotion intensity* to CT and ideals across different emotions.

Empirical study

Method

Two groups of graduate students at the University of Coimbra were used as participants in this study, for a total of 90 subjects.

One of the groups (n = 50) was presented with a long list of emotion-words (372) vertically spread over a booklet (rows) that also exhibited at its top the names of seven discrete emotions (columns). The instructions asked subjects to rate, using a 1-7 scale, "how suitable a given word-exemplar was to express a given emotion". Besides the names of the seven emotions (joy, fear, anger, sadness, disgust, surprise and love), an additional column entitled "not suitable to any emotion" allowed subjects to make that choice. There were no constraints imposed on the number of emotions to which a particular word could be attributed. The data collected this way were used to obtain two indices: (1) the "typicality-as-ideal" index, corresponding to the mean ratings obtained; (2) the "central tendency" index (CT), based on the matrixes of joint attribution of each pair of words to a same emotion category, irrespective of their ratings. Even if they come out of the same pool of data, these indices thus put in value different kinds of information, amenable to different kinds of treatments.

Subjects in the second group (n = 40) had to rate, in a similar 1-7 scale format, the intensity of emotion conveyed by each word-exemplar. Words were organized into separate sheets according to their respective emotion categories, with the name of the overarching category printed on top. Instructions explicitly required subjects to rate the intensity of the specified emotion as expressed by the words below it. These data were used to compute the intensity scores (mean ratings for each word within a category).

Results

Tables 1 to 3 summarize the functional relations of intensity to CT and "ideal" for six different emotions (two upper rows). In addition, the relations between"ideal" and CT can also be found at the bottom row in each table. CT was always plotted without accounting for the sign of the coordinates in the one-dimensional MDS solution: the horizontal axis in the graphs thus reflects absolute distance to CT maximum value ("zero" shifted towards left). Given that results for "Sadness" share the same pattern with all four emotions in tables 1 and 2 they are omitted here. The first major outcome to be noticed concerns the positive linear relation found between intensity and "typicality-as-ideal" across the four emotions in tables 1 and 2 (and also for "sadness"). That can be seen through looking at the middle row on both tables and checking the ANOVAs associated with the fit of the linear model (significant in all cases). This result converges with the findings reported by Horstmann using faces; moreover, since it is in no way restricted to "pure expressions" (an unattainable ideal when dealing with word materials) it goes one step further in tying up "intensity" to "typicality-as-ideal". On the basis of this outcome, it can be suggested that emotion expression stands in fact, overall, as an intensive dimension.

Table 1

Functional relations of Intensity Scores to CT and Ideal:
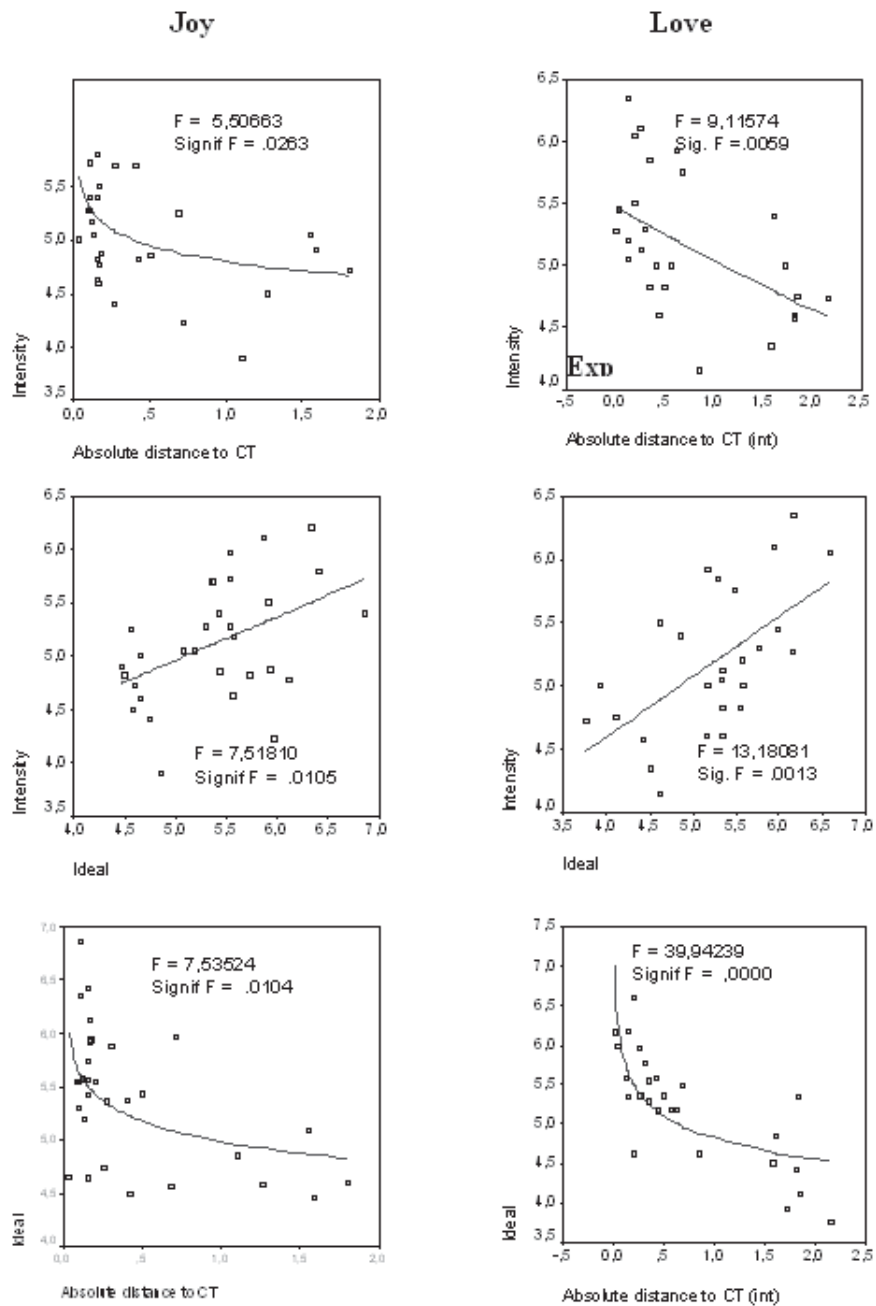*comparative patterns for Joy and Love*

Joy

F = 5,50663
Signif F = .0263

Love

F = 9,11574
Sig. F =.0059

F = 7,51810
Signif F = .0105

F = 13,18081
Sig. F = .0013

F = 7,53524
Signif F = .0104

F = 39,94239
Signif F = ,0000

Table 2

Functional relations of Intensity Scores to CT and Ideal:
comparative patterns for Fear and Anger
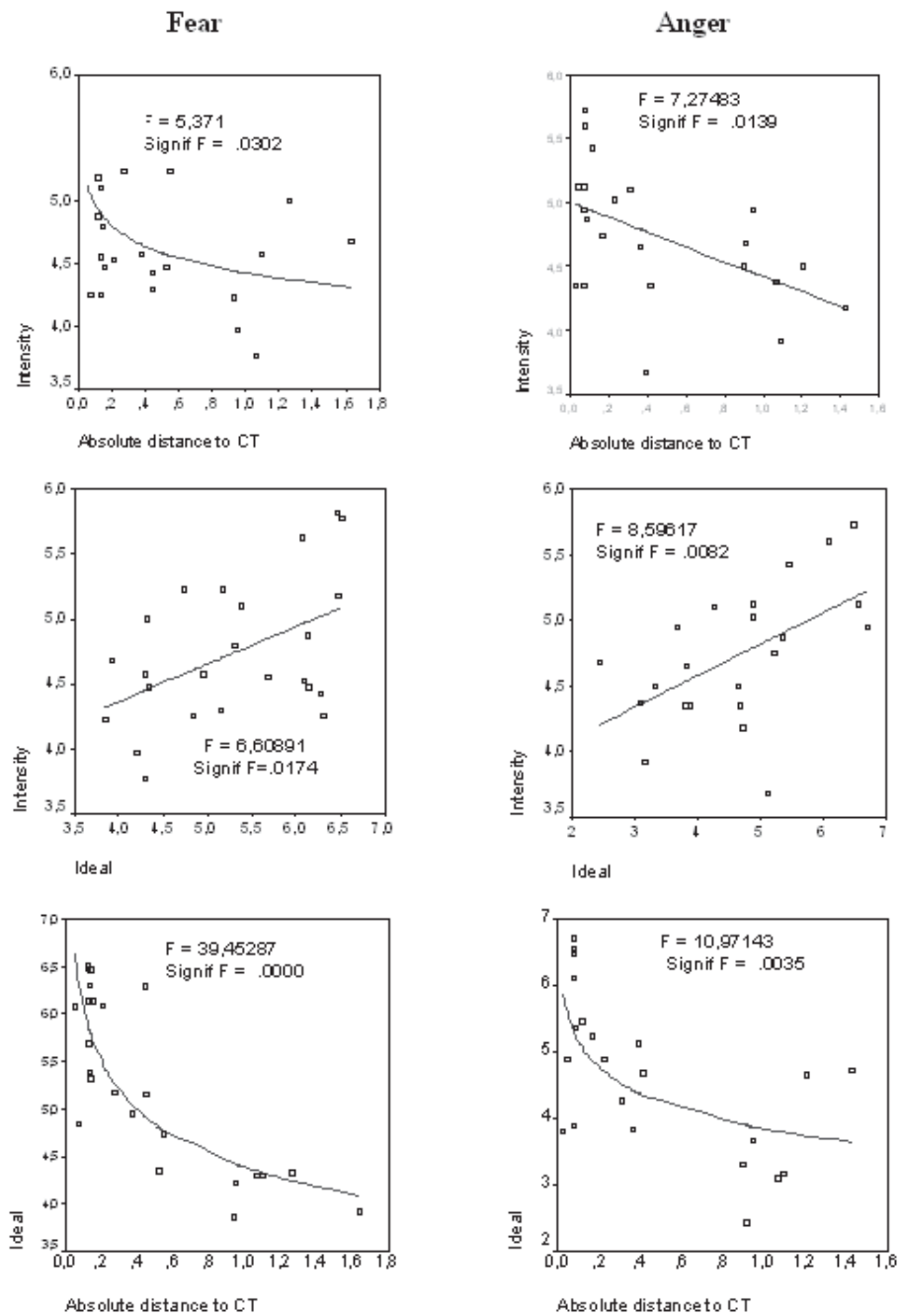
Table 3

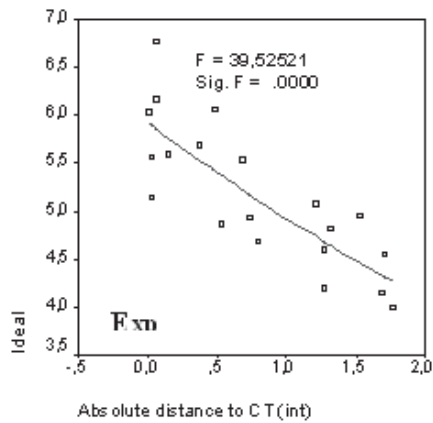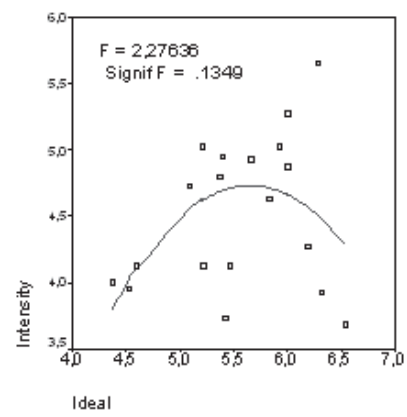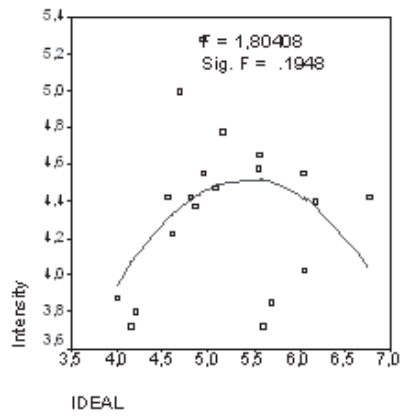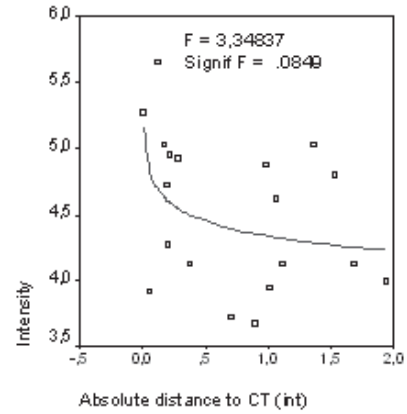*Functional relations of Intensity Scores to CT and Ideal:*
comparative patterns for Surprise and Disgust
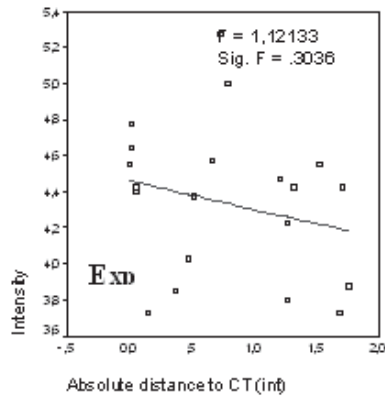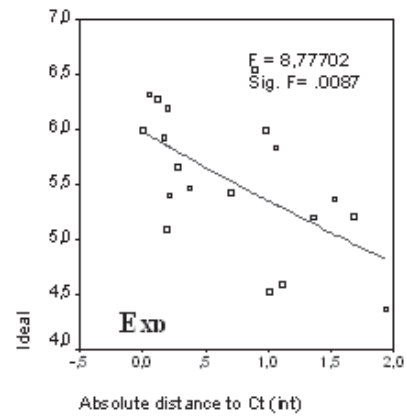
The second general outcome concerns the invariable declining pattern observed for "ideal" when plotted against "distance to CT". This trend has occasionally assumed an exponential, close to linear, form (such as in the bottom row of Table 3), but it usually exhibits a best power fit (Tables 1 and 2, third row). The characteristic shape of the functions concurs with the intradimensional similarity gradient observed by Shepard (1987) across multiple domains, which can be seen as an endorsement of the similarity measure used. The same can be said, in general, of the relation of "intensity" to "distance to CT", although the fits produced are somewhat less clear-cut (Tables 1 to 3, top row). On the whole, both ideal and intensity appear as closely related to CT, peaking at its maximum value («zero distance») and showing decreasing decline as distance to CT increases. This result argues against the generality of Horstmann's finding of a definite dissociation between CT and "ideal".

Finally, the third general outcome relates to emerging differences between emotions. Among the seven categories considered, surprise and disgust present a divergent pattern in regard to the other five, manifesting chiefly on the relations between""intensity" and "ideal". This is illustrated in the middle row of Table 3, where a quadratic relation provides the best fit to the data (contrasting with the more generally found linear fit). It should be noticed, nevertheless, that the fit is in neither case significant: more than revealing a true quadratic, U-shaped relation, it arises as a consequence of inconsistent behavior of intensity for the upper levels of "ideal", where it can show both increases and decreases in value. Regarding these two emotions, a relative dissociation of "intensity" and "ideal" can

thence be asserted – i.e., the goal of adequate emotional expression is not entirely in line with the growth of intensity of expression. As can be seen from the bottom row in Table 3, this doesn't disrupt the usual (decreasing) relation of "ideal" to "distance to CT": however, the previously found relation between "intensity" and "distance to CT" breaks down for disgust and surprise, ceasing to provide significant declining fits (see top row).

Final conclusions

As a general conclusion, it can be said that the overall intensive nature of "typicality-as-ideal" is clearly entailed by these results. There is, on this regard, a close agreement with Horstmann's findings. However, outcomes do not support the implication that "ideal" alone determines the graded structure of emotion categories. In fact, for all emotional categories envisioned, maximum ideal scores were located near the maximum value of "central tendency" («zero distance»). The prospect of deciding generally over the nature of the category of "emotion expressions" (taxonomic or goal-derived) by looking at its typicality determinants (CT or Ideal) thus seems to be called into question by these results. Furthermore, the relations of "intensity" to "ideal" were shown to vary with different emotions. It is therefore possible to conceive that ideals operate in different ways (e.g., more or less intensively) for different emotions, and engage by there, while determinants of typicality, into different relations with other kinds of variables. The bare uniformity of the notion of "ideal" may itself appear, under this light, as unwarranted overgeneralization. These accounted discrepancies with

Horstmann's results are plausibly stemming from differences in the stimuli materials as well as in the CT indices involved. On the whole, we take our CT index as more in accordance with the complexities of the notion of similarity than the usual average similarity score (Rips, 1989). As we see it, to obtain a CT measure freer from the salient perceptual/cognitive dimension standing for an "ideal" was a necessary methodological condition to allow for a less contrasted and more complex picture of the relations involving "CT", "ideal" and" "intensity". The interpretability of the one-dimensional solutions used for "CT" was only approached in broad heuristic terms. Nevertheless, for emotions such as surprise and joy, there seemed to exist simple satisfactory interpretations, resting on a bipolar "action impulse" dimension for the first, and on a bipolar "activation" dimension for the second. Other emotions, such as aversion, provided less clear pictures (even so, a mix of "bodily arousal" and "urgency for action" seemed to provide a reasonable interpretation). As stated before, the one-dimensional approach was just meant as an exploratory step towards alternative CT indices. To advance further within this logic, the dimensional solutions must be worked out first for each emotion, and the computation of "CT" made accordingly afterwards on the basis of the dimensionality found.

References

BARSALOU, L. (1983). Ad hoc categories. *Memory and Cognition, 11,* 211-227.

BARSALOU, L. (1985). Ideals, Central Tendency, and Frequency of Instantiation as Determinants of Graded Structure in Categories. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11(4),* 629-654.

BARSALOU, L. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101-140). Cambridge: Cambridge University Press.

BARSALOU, L. (1989). Intraconcept similarity and its implications for interconcept similarity. In S. Vosniadou and A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 76-121). Cambridge: Cambridge University Press

EKMAN, P. (1992). An argument for basic emotions. *Cognition and emotion, 6,* 169-200.

HORSTMANN, G. (2002). Facial expressions of emotion: Does the Prototype represent Central Tendency, Frequency of Instantiation, or an Ideal? Emotion, *2(3)* ,297-305.

RIPS, L. (1989). Similarity, typicality and categorization. In S. Vosniadou and A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 19-75). Cambridge: Cambridge University Press

ROSCH, E. (1973). On the internal structure of perceptual and semantic categories. In T. Moore (Ed.), *Cognitive development and acquisition of language.* New York: Academic Press.

ROSCH , E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192-233.

ROSCH, E. & MERVIS, C. (1975). Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, *7,* 573-605.

SHEPARD, R. (1987). Towards a universal law of generalization for psychological science. *Science, 237,* 1317-1323.