Are discrete emotions goal-derived or taxonomic categories? A study based on lexicon[φ].

Armando M. Oliveira[*]
Marta Teixeira*
Eduardo R. Santos*
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 linear relation between intensity ratings and "typicality-as-ideal" in all cases, and to a differentiated pattern of relations between both and "central tendency" (indexed on the basis of a multidimensional scaling measure), varying across emotion categories.

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.

Tipicality

In its most general sense, stemming from Rosh pioneering studies (1973; 1975), typicality refers to a graded structure underlying category membership as well as 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 incline 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;

[φ] Work supported by the project POCTI/PSI/41235/2001, financed by the Portuguese Foundation for Science and Technology.

[*] Institute of Cognitive Psychology, University of Coimbra, Portugal – *l.dinis@fpce.uc.pt*
[**] University of Lisbon.

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 on which typicality judgments are based upon.

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: we will call it, more specifically, "representativeness" instructions. As will become clear below, this also closely matches 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 the other members of the category (Barsalou, 1983).

Horstmann has embraced this operational understanding of family resemblance as a distance to CT in his paper. Given that the stimuli used in his study were drawings of five facial expressions varying across five levels of intensity, CT scores were obtained by having participants to judge the similarity of all pairs of faces within each emotion category: the average similarity of each face to all the other category members was then taken as an index of CT (Horstmann, 2002, 299-300). This is just one possible way of acknowledging the functional equivalence between average similarity to members and similarity to their 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 that category by a number of subjects (n = 66), an input matrix of proximities was built up (i.e., one for each emotion category). This matrix, interpreted as a matrix of similarity between word pairs, was then treated through *multidimensional scaling* (Proxcal) constrained to a single dimension. 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 it, 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 lying behind is 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 correspondence 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, the goal of our study was not to find the source of typicality in the domain of emotion expressions, but to explicitly address the relations of "tipicality as ideal" and CT with *intensity* as a major dimension of felt and expressed emotions.

Besides "ideal" and CT scores, a "frequency measure" was also collected. Following Barsalou, it is important to distinguish between "familiarity", which is a category-independent measure, and "frequency of instantiation" (FI), applying to how often someone experiences an entity as an instance of a category (Barsalou, 1985, 631). In his work, Horstmann resorted to overt questioning about FI (2002, 299). The particular "frequency" measure we used was the number of times one word exemplar was attributed to a given emotion-category. This index shares with FI the property of being category-specific. However, it doesn't correspond to a specific judgment over frequency. Given that there may exist large differences between people estimates about an exemplar's FI and the number of times it is attributed to the associated category, these two measures are hardly comparable; for that reason, no account will be given here of the results found for "frequency".

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 as to the determinants of typicality being 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 usual 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 associated with categories, 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 happens with 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 study (2002)

The study by Horstmann embodies a different logic from those of Barsalou, who aimed at demonstrating that ideals can determine graded structure; for that, 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), which may qualify as a common taxonomic as well as a goal-derived category in light of the above-presented guidelines. The issue at stake was whether, on the basis of the observed determinants of typicality (namely "ideals" or CT), a decision could be made regarding the 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 each of the other exemplars in the category. Given the perceptive salience of intensity in the particular materials used – schematic drawings of facial expressions varying 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, made possible by resorting to other emotion-expressive materials, could furnish different results. A different way of computing the CT index, in the spirit of "mental maps" embodied by MDS techniques, might also be helpful on this regard.

Also, as Horstmann himself acknowledges, his stimuli materials are entirely made up with 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 results in the domain of emotion expressions (not just facial ones). More centrally, however, its goal shifts away from the issue of typicality to specifically address the relations of *intensity*, as a central dimension of felt and expressed emotions, 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 111 subjects.

One of the groups (n = 66) was presented with a long list of emotion-words vertically spread over a booklet (rows) that also contained at its top the names of seven discrete emotions (columns). The instructions asked subjects to rate, using a 1-6 scale, "how suitable a given word-exemplar was to express a given emotion". Besides the names of the seven emotions (happiness, 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 as to the number of emotions to which a given word could be attributed. The data collected this way were used for obtaining two indices: the "typicality-as-ideal" index, on the basis of the ratings made (mean ratings), and the CT index, based on the matrixes of joint attribution of each pair of words to a

same emotional category, irrespective of their ratings. Even if they come out of the same pool of data, these indexes put in value different kinds of information, amenable to different kinds of treatments.

Subjects in the second group (n = 45) had to rate, in a similar 1-6 format scale, 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 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 and 2 summarize the functional relations of intensity to CT and Ideal indices for two pairs of emotions (Happiness, Sadness; Fear, Anger). The two top rows on each table correspond to different ways of plotting CT in the horizontal axis - either accounting for the sign of the coordinates (upper row) or alternatively reflecting their absolute distance to the center ("zero" shifted towards left). The bipolarity allowed for in the first case can be helpful as a check on the interpretability of the one-dimensional solution, while the absolute distance to CT offers a more comparable picture with the common CT measures. The patterns exhibited by emotions in the left column (Happiness and Fear) are quite similar in both tables, and differ from the ones presented on right. Given that results for "Aversion" and "Surprise" share the same pattern with Happiness and Fear they were omitted here.

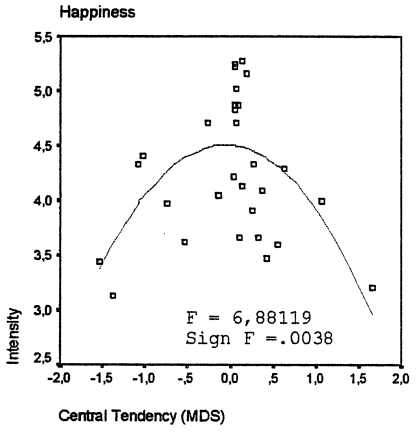
The first major outcome to be noticed is the invariable linear relation between intensity and "typicality-as-ideal" across all emotions. That can be seen through looking at the bottom row on both tables and checking the ANOVAs associated with the fit of the linear model (highly significant in all cases). 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 is by nature an intensive dimension; it also seems sound to expect that the use of blends of facial expressions in forthcoming studies will not significantly alter this picture (which remains to be seen).

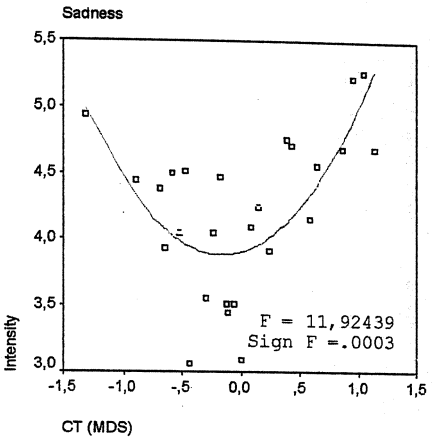
The second general outcome concerns the differences occurring between emotions. While the relation between intensity and ideal favors uniform behavior across emotional categories, things go differently when it comes to the relation between intensity and CT. This is best illustrated by the second row in both tables, were two opposite patterns can be observed. While for Happiness and Fear (left column, tables 1 and 2) intensity decreases dramatically with absolute distance to CT (according to a power best fit), for Sadness and Anger, on the contrary, it shows a linear increase. This result points towards a twofold nature of emotional categorization. In some cases, CT (indexed by "zero" on the abscissa) appears to be closely related with intensity and ideal; on some others, it strongly diverges from both. The top row in Table 1 makes that clear through the symmetrical quadratic adjustments (concave downward for Happiness, an upward for Sadness). On Table 2, Anger shows a descending linear pattern, which can be looked upon as just another of way of making intensity diverge from CT (in fact, a slightly quadratic fit proved better than the linear one, but depended on a single right extreme point)

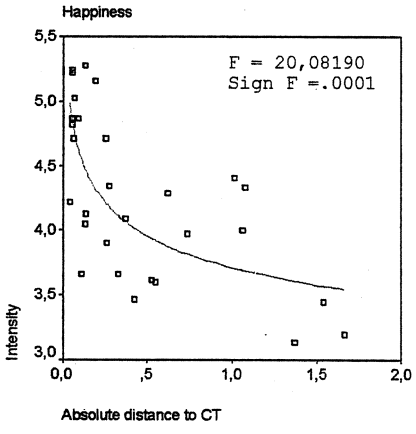
Table 1

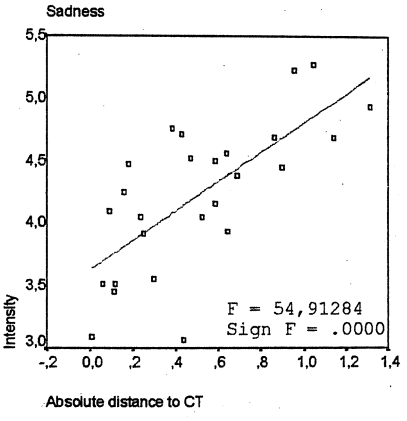
Functional relations of Intensity Scores to CT an Ideal: comparative patterns for Happiness and Sadness

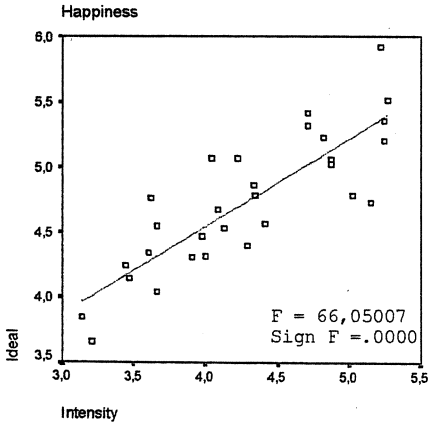
Happiness | **Sadness**











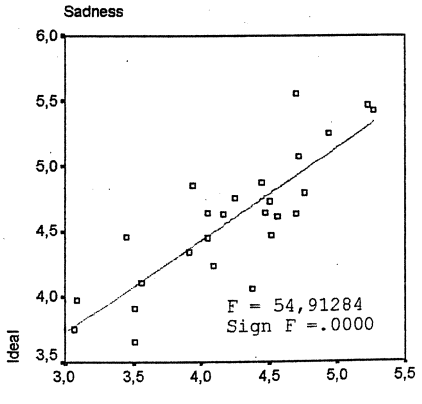


Table 2

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

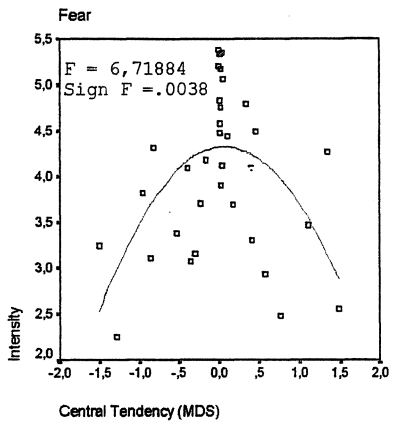
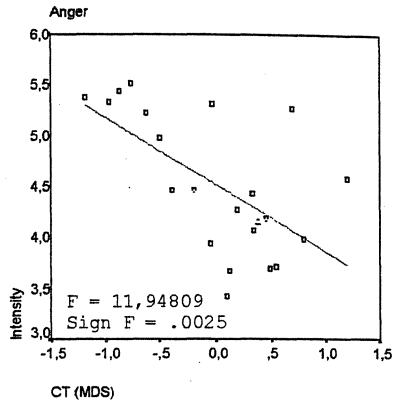
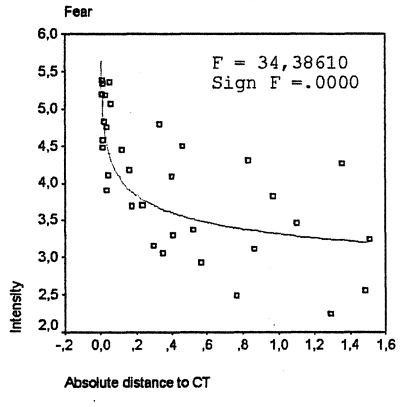
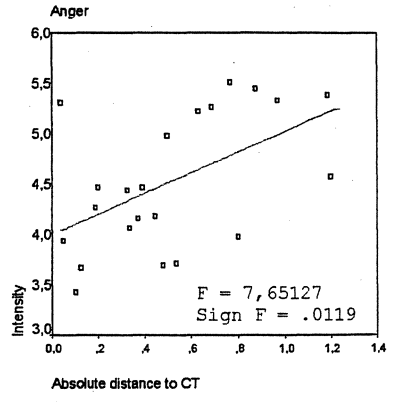
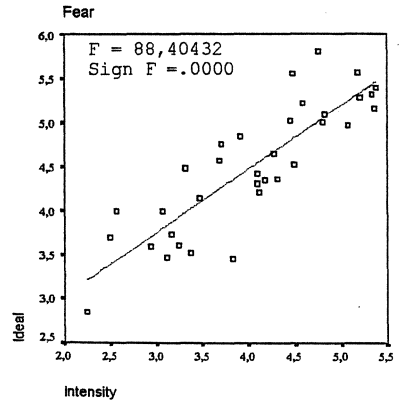
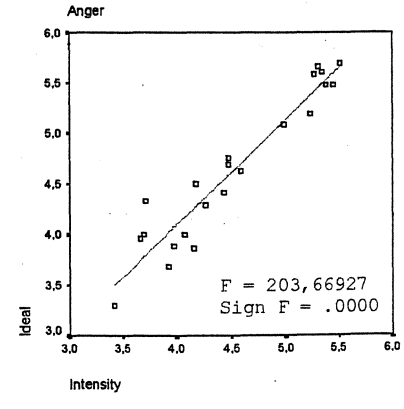
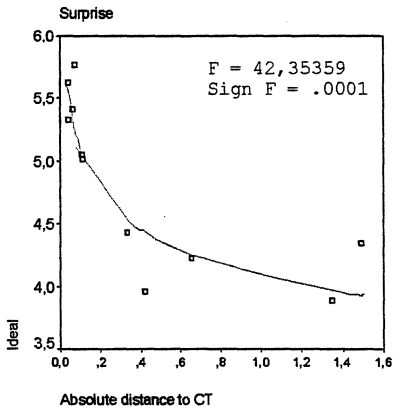
Fear	**Anger**
	
	
	

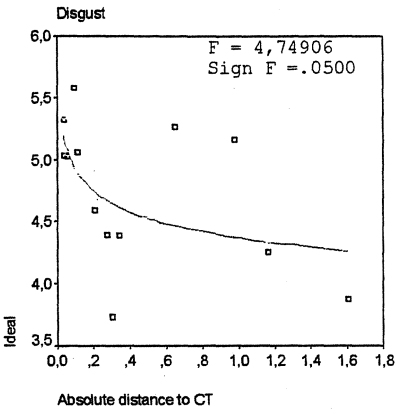
Table 3

Functional relations between CT (absolute distance) and Ideal: comparative patterns for all six emotions

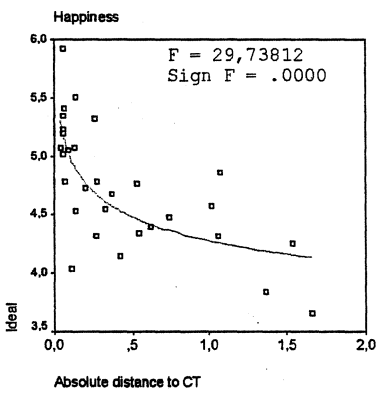
Surprise



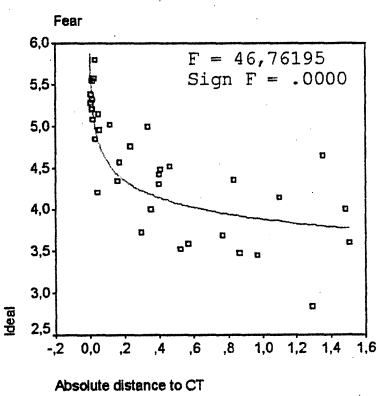
Disgust



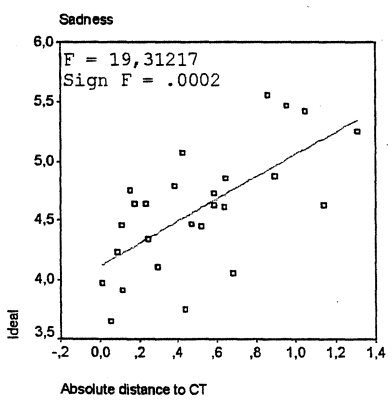
Happiness



Fear



Sadness



Anger

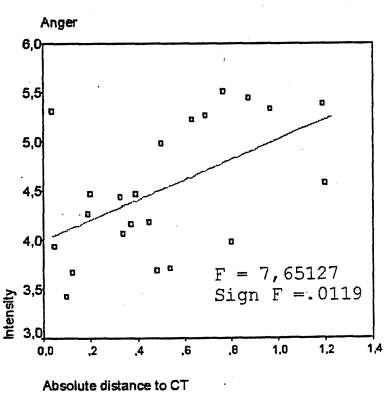


Table 3 presents the functional patterns relating "typicality-as-ideal" to absolute distance from CT for all the specific emotions considered. As could be anticipated from the prevailing linear bond between intensity and ideal, all similarities and differences ascertained were replicated by replacing intensity for "ideal" on the ordinate. The contrast between the bottom row and the two upper rows on the table clearly illustrates the previously found distinction among Sadness and Anger, on one side, and the other four emotions. Furthermore, for the majority of emotions, the same best power fit observed with intensity could also be reproduced with "ideal" on the ordinate.

Final conclusions

As a general conclusion, it can be said that the intensive nature of "typicality-as-ideal" is clearly entailed by these results. There is, on this regard, a close agreement with Horstmann findings. However, they do not support the implication that ideal alone determines the graded structure of emotion expression categories. In fact, four of the emotional categories envisioned showed maximum intensity and maximum ideal scores near the "central tendency", while two of them exhibited a clear dissociation between "ideal" and "central tendency". The role played by CT seems therefore to vary with different emotions, leaving aside the possibility of clearly deciding whether emotion expressions are structured like common taxonomic or goal-derived categories.

This disagreement with Horstmann results is plausibly due, on the one side, to differences in the stimuli materials and, on the other side, to the CT indices employed. On the whole, we take our CT index, obtained through an MDS solution, as more in accordance with the complexities of the notion of similarity than the usual average similarity score (Rips, 1989). By making possible to obtain a CT measure freer from the salient perceptual or cognitive dimension that stands for an "ideal", a more complex picture of the joint role of CT and "ideal-intensity" can emerge. In particular, it can be conceived that intensive aspects concerned with ideals must be combined with more "taxonomic" ones, dependent on different dimensions, to support emotional categorizing.

The interpretability of the one-dimensional solution used for CT was only approached in broad heuristic terms. Nevertheless, for Sadness and Anger there seemed to exist simple satisfactory interpretations, resting on a "passivity-activity" dimension (spleen-despair) for the first, and on a "high-low intensity" dimension for the second one (in fury-annoyed). The other four emotions provided less clear pictures (however, in most cases, the use of bi-dimensional solutions offered plausible interpretations). As stated before, the one-dimensional approach was meant as a tentative exploratory step to alternative CT indexes. To advance further within this logic the dimensional solutions must be worked out first for each emotion and the computation of CT made afterwards in accordance.

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.