Acta Psychologica 138 (2011) 405–418

Contents lists available at SciVerse ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Perceptual ratings of opposite spatial properties: Do they lie on the same dimension?

Ivana Bianchi [a,*], Ugo Savardi [b], Roberto Burro [b]

[a] Department of Educational Sciences, University of Macerata, P.le Bertelli 1, 62100 Macerata, Italy
[b] Department of Philosophy, Education and Psychology, University of Verona, Lungadige Porta Vittoria 17, 37129 Verona, Italy

ARTICLE INFO

Article history:
Received 29 December 2010
Received in revised form 23 August 2011
Accepted 26 August 2011
Available online 5 October 2011

PsycINFO classification:
2323
2340

Keywords:
Perception and cognition of spatial opposites
Unidimensionality
Experimental phenomenology
Phenomenological psychophysics
Factor analysis and Rasch models applied to perceptual ratings

ABSTRACT

The issue of unidimensionality is dealt with in various research areas in the field of Psychology (e.g. conceptual spaces, semantic modeling, psychometrics) and always involves spatial modeling. An investigation of the dimensionality of opposite spatial scales (even basic) has however not yet been carried out. In this paper we look at whether opposite judgments of height, size, width and length (*high/low*, *large/small*, *wide/narrow*, and *long/short*) imply underlying unidimensional continua. In three experiments, independent ratings for the 8 above mentioned properties were elicited with participants looking at photographic representations of various objects (Study 1), real life objects (Study 2) and spatial extensions in object-independent conditions (Study 3). Explorative and confirmative factor analysis and Andrich Extended Rating Scale Models were applied in order to determine whether the ratings referred to opposite scales on the same linear continuum. Results from the three studies consistently revealed that this is not the case. A joint analysis of the data showed interesting interactions between the spatial properties analyzed suggesting a possible explanation for the lack of unidimensionality.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

1.1. The relationship between opposites and dimensions

In the field of Psychology, what lies between opposites is generally referred to as a *dimension*. Dimensions represent the psychological counterparts of physical continua. They are essential to the understanding of conceptual spaces, since dimensions are conceived to be the axes on which conceptual spaces are defined (e.g. Gärdenfors, 2000, 2007; Hollins, Faldozski, Rao, & Young, 1993; Miller, 1996; Picard, Dacremont, Valentin, & Giboreau, 2003). "Some theorists think of attributes as dimensions. In that case, the N attributes of a given referent define an abstract N-dimensional space (a hyperspace); any particular instance can be thought of as a point in that space, its location given by its value on each dimension. (…) The size (…) of the cup, for example, can be *large or small* (modifiers), its height (…) can be *tall or short* (modifiers), and so on. Each modifier specifies the position of a referent along a particular dimension; the full set of modifiers specifies a particular location in the N-dimensional space" (Miller, 1996, p. 193).

If dimensions are the cognitive structures which account for the psychological variations of a property from *x* to its opposite and if these structures are the foundation of conceptual spaces, one might expect the investigation of the structure of dimensions to be a central issue in cognitive psychology. In contrast, to date, a systematic experimental analysis of the latent structures underlying judgments referring to opposite properties has still to be developed. For example, we do not have any empirical evidence that there is a single dimension underlying ratings of 'large' and 'small'. This kind of question is familiar in areas of psychological research such as psychometrics as applied to the modeling of emotional and social dimensions (see next section) but it is not explored as a general cognitive phenomenon as regards the structure of opposites. Neither is it explored in terms of those opposites which pertain to *spatial* perceptual dimensions and which are implicit in the concept of dimension, as clearly emphasized in Miller's quotation above. This is even more clear in Garderfors' geometrical model of conceptual spaces (2000, 2007).

1.2. The unidimensionality of opposites: A methodological or cognitive question?

The idea that two contrary properties lie on the same continuum has formed the basis of various methodologies used in experimental research in the field of Psychology since the late sixties (e.g. the differential semantic method, Likert scales, etc.); however, an increasing number of methodologists are facing problems connected to the assumption of unidimensionality in opposite scales (Conrad et al., 2004; Idaszak & Drasgow, 1987; Millis & Neimeyer, 1990; Podsakoff,

* Corresponding author. Tel.: +39 0733 258 4320.
E-mail address: ivana.bianchi@unimc.it (I. Bianchi).

0001-6918/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2011.08.003

MacKenzie, Lee, & Podsakoff, 2003; Yorke, 2001). Researchers have proved that using opposite wording or reverse-coded items (e.g. items asking for ratings of *comfort* versus *stress*) affects participants' responses (Billet & McClendon, 2000; Chiorri, Anselmi, & Robusto, 2009; Harris & Bladen, 1994; McGee, Ferguson, & Seers, 1989; Tracy & Johnson, 1981; Rodebaugh, Woods, & Heimberg, 2007). The effect of inverse wording has usually been addressed as a methodological problem (cf. Barnette, 2000; Hinkin, 1995; Idaszak & Drasgow, 1987; Podsakoff et al., 2003) or related to social desirability or acquiescence response biases (Feather, 1980; Lorenzo-Seva & Rodrigues-Fornells, 2006; Ray, 1983, 1985). However, it may be that these differences are not simply caused by this type of factors, but may also reflect the fact that the opposite properties studied do not lie on the same continuum.

While searching for an empirical validation of their theories and models, psychologists studying emotions, personality and social traits addressed the question of whether the traits that they were proposing (e.g. *pleasant–unpleasant*, *relaxed–tense*, *happy–depressed*) were opposite poles of a single dimension or rather referred to independent constructs (for a review, see Russell & Carroll, 1999; Schimmack, 2001). For instance, it has been shown that *masculine–feminine* (Marsh & Myers, 1986) and *optimism–pessimism* (Kubzansky, Kubzansky, & Maselko, 2004; Pinquart, Fröhlich, & Silbereisen, 2007) behave as distinct scales rather than single unidimensional constructs, whereas *individualism–collectivism* (Freeman & Bordia, 2001), *burnout–engagement* (González-Romá, Schaufeli, Bakker, & Lloret, 2006), *pleasantness–unpleasantness*, and *happiness–sadness* (Watson & Tellegen, 1999) behave as opposite poles of a single underlying dimension. Is this only the case with social and emotional dimensions or do these data indicate a more general phenomenon regarding the cognitive structure of opposites? If one looks at how the issue of unidimensionality in contrasting properties has been dealt with in cognitive research, one first of all has to acknowledge that the problem has not been focused on as closely as in other areas and that in any case deeper experimental investigation of this and related issues is required.

1.3. What we have learned from the study of opposites in cognitive linguistics

The structure of antonyms and their corresponding dimensions have been a central issue in linguistics and cognitive semantics (Croft & Cruse, 2004; Fellbaum, 1995; Jones, 2002; Muehleisen, 1997; Murphy, 2003; Paradis, 2001, 2008). One of the frequently emphasized paradoxes of antonyms is that they are in complete contrast and yet they also appear to have something in common. This characteristic of being simultaneously close and distant has led to the conclusion that two antonyms are the endpoints of an underlying continuum, or, in other words, that "antonyms name opposite sections of a single scale" (Lehrer & Lehrer, 1982, p. 484; see also Cruse, 1986). Since the early seventies, this has been the idea underlying concepts of semantic dimension and markedness (Clark, 1973). In any pair of opposites one of the two properties is expected to be the uncommitted (or unmarked) pole, while the other represents the committed (or marked) pole. A property is said to be committed if it implies a particular value when used in a question, and impartial or uncommitted if it does not. For example, *old* is uncommitted in a question like "How *old* is Maria?", given that this question can be used whether or not one knows Maria's age, i.e. whether Maria is young, old or middle aged. In contrast, a speaker would only ask "How *young* is Maria?" if there is some reason to believe that she is a young woman. The adjective *young* is therefore committed. This might seem to be simply a pragmatic rule, but has been taken to be a fundamental cognitive principle: the uncommitted pole covers all the possible variations of a given quality and this proves the existence of a cognitive continuum (i.e. a dimension) to which the adjective applies. For example, the dimension "length" is derived from (and in this case also morphologically related to) *long*: in the pair *long–short*, *long* is the uncommitted, unmarked pole (for a critical overview on the debate on markedness, see Haspelmath, 2006).

The idea that opposites presuppose an underlying continuum has been a default assumption in linguistics and cognitive semantics for more than two decades even though it has been noted that in everyday language people describe their perceptions in terms of opposites ("you are driving fast", "the walk is long", "the room is small", "it's hot today") instead of using unidimensional scales (respectively, velocity, size and temperature). Mosconi (1967) suggested that in natural language (or, as he said, in the "phenomenal" use of language) people shift toward the use of unidimensional scales only when certain processes are involved: a) when measuring: observers switch from "this cable is very *short*" to "the cable is 2 cm *long*" and b) when making comparisons: observers switch from the description of two very *short* pieces of cable of 2 and 3 cm to the phrase "this (3 cm) is *longer* than that (2 cm)". In all other cases, when spontaneously describing their direct perceptual experiences, people refer to opposites.

Cruse and Togia (1995), while searching for cognitive models of antonymy, drew attention to the difference between opposites presupposing a monoscalar system (such as *long–short*) and opposites presupposing a biscalar system (such as *hot–cold*). In the former type, opposites form a single scale, "calibrated in terms of everyday conventional units" (p. 115). Conversely, opposites of the second type "intuitively involve two distinct gradable properties (…). The properties are typically not calibrated, or, at least, it is not this mode of conceptualization which governs the behavior of the antonyms (…). They are characterized by scales arranged end-to-end and with adjacent zeroes." (p. 115). The authors based their intuitions on descriptions of linguistic typical behaviors, but also on what these terms intuitively involve. They also occasionally made reference to sensorial experiences ("pouring boiling water into cold water makes it first less and less cold, then tepid, then gradually hotter", p. 116).

2. Experimental phenomenology applied to the study of the structure of dimensions

"When the dimensions are seen as cognitive entities (…) their structure should (…) be determined by *psychophysical* measurements that determine how our concepts are represented" (Gärdenfors, 2000, p. 3–4). In this paper we address the issue of unidimensionality by focusing on the opposite perceptual scalings of the basic spatial properties pertaining to extension. We used an experimental phenomenology approach (Bozzi, 1989; Gepshtein, 2010; Kanizsa, 1979; Kubovy, 2003; Kubovy & Gepshtein, 2003; Pinna, 2010; Schmicking & Gallagher, 2010), or what has been defined as phenomenological psychophysics (Kubovy, 2003; Kubovy & Gepshtein, 2003; Burro, 2009), by matching the scaling of a set of objects expressing the perceived degree of, for example, *length* to the scaling of the same objects according to the perceived degree of *shortness*. Instead of matching a physical scaling (e.g. length in cm) and a perceptual scaling (e.g. perceived degree of length) as in traditional psychophysics, we matched pairs of scalings obtained from subjective ratings of the following qualities: *high* and *low*, *large* and *small*, *wide* and *narrow*, and *long* and *short*. The application of Rasch models allowed us to compare these scalings after the subjective variability component was removed from the data. This meant focusing on the δ_i parameters (see the discussion of the β_n and δ_i parameters in the family of Rasch' models in the introduction to the studies, Section 3.1). This was an interesting approach, because it enabled us to focus on the characteristics of the "items" (in our case, on the description of the stimuli in terms of the perceptual presence of the properties under observation) rather than on information about observers' ability to detect a certain property.

The scalings were obtained with reference to pictures of ecological objects (Study 1), direct perception of the same objects (Study 2) and direct perception of spatial environmental extensions (Study 3). We considered that if the description of the spatial characteristics of

these stimuli in terms of the two contrasting properties did not provide evidence of a common underlying dimension, this would shed new light on the issue of unidimensionality, showing that the issue already exists when dimensions related to the extensional experience of space are considered.

Since we are studying the dimension underlying two scalings, it might be of help to think of this research in terms of an investigation into 'conceptual categories' of space rather than spatial perception per se. However, we would like to emphasize that in the studies presented here we match scalings which in the first study referred to pictures of real-life objects and in Studies 2 and 3 were direct estimates of real-life objects and spatial extensions under strict observation. Therefore we are studying unidimensionality not merely on a conceptual (or linguistic) level, but grounding it on the structure of scales describing quantifications of perceived properties. This research, in a more general sense, adds to the studies on the perceptual foundation of cognition (for an overview see Pecher & Zwaan, 2005). More specifically, it contributes toward a better understanding of the structure of opposites in direct experience developed within the field of experimental phenomenology of perception (for an overview, see Bianchi & Savardi, 2008a,b). Results from previous studies (Bianchi, Savardi, & Kubovy, 2011) indicate that in order to characterize dimensions and their poles based on their phenomenal structure, we need two types of information (see Table 1): a) metric information about the asymmetry of the poles and the width of the neither-nor region and b) topological information about each pole (is it a bounded interval, an unbounded interval or a point?) and the neither-nor region (are intermediates nonexistent, an interval or a point?). Using this metric and topological information, the properties studied in this paper (*high*/*low*, *large*/*small*, *wide*/*narrow*, and *long*/*short)* which describe the extension of objects and environments along three main spatial axes (sagittal, coronal and gravitational) can be described in terms of gradations of poles and intermediates (Table 1: URB).

In traditional linguistic classifications they are typical representatives of the category of gradable adjectives. In recent linguistic analyses they are still conceptualized as unidimensional: they presuppose a monoscalar system (Cruse & Togia, 1995) and "crucially make use of the same dimension and degrees (e.g. both tall and short map their arguments onto corresponding degrees of height) but express inverse ordering relations" (Kennedy & McNally, 2005, p. 351). The point that we raise in this paper is that it might be an error to generalize this claim directly from linguistics (even cognitive linguistics) to cognition grounded in perception and indeed the story may be more complex and the conclusion different if the dimensions emerge from underlying perceptual scalings.

Table 1
Different types of opposites, defined by metric properties and refined by topological properties. Table adapted from Bianchi et al., 2011.

Type	Example	Metric information		Topological description
		Asymmetry of the poles	Extension of the intermediates (proportion on total dimension)	(Pole A, neither-nor, pole B)
I	Incomplete–complete	Strong	0	UNP
II	Large–small	Moderate	1/3	URB
III	Full–empty	Moderate	2/3	PRP, BRB
IV	Inside–outside	Minimal	0	BPB, PNP, UPU

Note: In the topological description, we coded responses to the two poles as P (point), B (bounded range), or U (unbounded range) and the responses to the neither-nor region as N (none), P (point), or R (range).

3. The studies

In the three studies presented here, we verified if the scalings of given sets of stimuli obtained from judgments made by participants in terms of one property or its opposite imply that there is a single common linear dimension underlying the two series of judgments. In the first experiment, participants looked at photographs of various different objects and were asked to estimate the degree to which they showed eight spatial properties (*high*/*low*, *large*/*small*, *wide*/*narrow*, and *long*/*short*). In the second experiment, the same task was done looking at real life objects. In the third experiment, the estimates referred to variations of spatial extensions in object-independent conditions.

Comparing ratings can be a delicate matter, since problems of units of measurements, the origin of the scales and their lack of linearity have to be taken into account. Traditional unidimensional scalings only partially deal with these issues.

3.1. A methodological note on the studies: The psychophysical application of Item Response Theory models

The dimensionality of opposite scales is usually studied using factor analysis or Rasch models (Rasch, 1960; ed. 1980) and their extended versions (Andrich, 1978a,b,c, 1988). In our studies we applied both methodologies. Factor analyses, although more familiar, are based on a covariance matrix, which is known to be affected by distribution artifacts. Rasch analyses are not biased by item distributional forms. They are normally used in the context of Item Response Theory to measure abilities, attitudes or personality traits. They have however also been used to study cognitive processes such as conditional reasoning (Bonnefon, Eid, Vautier, & Jmel, 2008), semantic categorization (Verheyen, Hampton, & Storms, 2010) and in psychophysical research (Vidotto, Robusto, & Zambianchi, 1996; Burro, Sartori, & Vidotto, 2011). A few introductory notes may be useful to clarify the logic supporting Rasch models and to explain the data analyses conducted in our three studies.

Rasch's simple logistic model is the basis of Item Response Theory models for the transformation of ordinal observations into linear measures. It expresses, according to logistic distribution, the probability of good/bad performance as a function of the participant's ability and the difficulty of the items. The more able the participant and the less difficult the item, the greater the probability of good performance (with dichotomous data) or the probability of participant assigning an item to a given category (with category ratings) — see Appendix 1 for further details.

Let us now focus on three aspects of these models that we consider to be particularly suitable:

1) *Participants* versus *stimuli*. Rasch models allow the researcher: *i*) to measure and rank participants based on their ability to perceive a target property (β_n values) — in our research the target properties were *high*, *low*, *large*, *small*, *short*, *long*, *wide*, and *narrow*; *ii*) to rank a set of stimuli according to the degree to which they manifest a target property (δ_i values).[1] Since we are interested in scaling the set of stimuli, Rasch models will enable us to concentrate on this (δ_i). In saying that we focus on the scalings of the stimuli, we are not saying that we are interested in analyzing the physical properties of stimuli, but in their phenomenological characteristics, i.e. how they appear to the observer. In effect, the perception of objects as being *low* or *high* is a genuine psychological fact, as is the perception of movements as either *fast* or *slow*. Indeed, in terms of physics, objects are not *low* or *high* but extend

[1] Traditional applications of Rasch models relate to ability tests and thus they use a language that refers to "able" or "not able" respondents, "easy" or "difficult" items. In application to perceptual scales, this language needs to be modified.

x centimeters and movements are not *fast* or *slow*, but cover *x* meters per second. Therefore, studying the scalings of objects based on the degree of "highness" and "lowness" which is perceived means measuring a psychological dimension. Part of the phenomenal experience of these properties is their objectiveness, i.e. the fact that they are perceived as properties *of* the objects. In this sense, Rasch models (and in general the theory of conjoint measurement) are suitable in terms of distinguishing between two psychological measurements, the phenomenological characteristics of the stimuli and the ability of the subject to recognize them.

2) *The β_n (participants) and δ_i (stimuli) scales have two important characteristics*. First, object/stimuli and participants are respectively calibrated and ranked on a single continuum following an additive criterion. It is assumed that participants who succeed in giving correct responses to more difficult stimuli (i.e. those occupying a high ranking on the continuum) will also be able to respond to those requiring less ability. For instance, when evaluating how *low* objects are, participants will differ from each other in their ability to discriminate this property. Each participant is then represented as a point along a linear continuum, with other people occupying various positions on the same continuum, depending on their ability. Similarly, the resulting order of the stimuli represents the degree to which the property *low* is manifested. The stimuli which best represent the target property would appear on one half of the continuum; those which display the property to a lesser degree would appear on the other half of the same continuum.

   Second, the continuum along which stimuli/items and participants are scaled is a logit scale. This means that the values obtained from raw data (which will be used to verify the unidimensionality of the two opposites scales) are not affected by the distortion of the distances at the extremes of the scale which characterizes raw scores obtained using category scales (Wright & Masters, 1982). Moreover, since we are using the same units of measurement for all the target properties, we have the chance to compare different scalings directly. This is a very important requisite for our research, since we need to match two scalings, one produced when describing a property (e.g. *high*), the other produced when describing the opposite property (e.g. *low*).

3) *Model-data fit and separation indices*. Rasch models are based on two assumptions which are partially implied by the cumulative nature of the scales: first, the higher the ability of the participant, the higher the probability that he/she will be able to discriminate more items/stimuli than a participant with lower ability; second, participants will give a better discriminative performance when presented with an easy stimulus than with a difficult one. Because of the probabilistic nature of the model, it is conceded that empirical data might deviate from that predicted by the model. Various measures of fit are available (Andersen, 1982; Gustafsson, 1980; Molenaar, 1983; Niemoller & van Schuur, 1983; Rost, 1982; van den Wollenberg, 1982; Wright, 1980). If the model-data fit turns out to be good, we can exclude the possibility that, when matching two scalings, the absence of unidimensionality is due to: a) idiosyncratic responses of a small subset of participants; b) the fact that participants have changed their judgments across trials, depending on which comparison objects are used when judging various different stimuli; c) the idiosyncratic behavior of some of the stimuli.

   In addition to the indices of fit, two further indices are provided by these models: the Person separation index and the Item separation index (Wright & Masters, 1982). These express the ratio between the adjusted person standard deviation (or the adjusted item standard deviation) and the mean standard error. In terms of the corresponding reliability coefficients (Wright & Stone, 1979, pp. 162–165), when the values of these indices approach 1, this means that the set of stimuli (and the sample of participants) is homogeneously distributed to cover the whole continuum. This is also relevant to our study. We took two steps when analyzing data. Firstly, for each target property, we verified whether the stimuli were distributed along a continuum (this was done by evaluating the fit with the model) and whether they were well enough distributed to cover the whole continuum (information provided by the Item separation index — or reliability separation coefficient). This provided the guarantees described above in points a, b, c. Secondly, the two scalings referring to the two opposite properties were compared in order to verify if the two continua were in fact the same.

4. Study 1

The aim of the study was to verify whether, when people estimate for example how *high* and how *low* objects are, they are judging the same characteristic and making estimates of the same spatial construct. The same went for all the 4 pairs of properties we studied. We conducted the analysis using a set of ecological objects.

4.1. Method

4.1.1. Participants

179 undergraduate students at the University of Verona (113 females and 66 males, mean age 20.9).

4.1.2. Procedure

The study was conducted in a classroom at the University of Verona. Each participant received two A4 sheets of paper showing twenty-four photographic representations of objects (size 6×6 cm; gray scale) and eight A4 response sheets, each related to one of eight spatial properties: *high*, *low*, *large*, *small*, *wide*, *narrow*, *long* and *short*. The photos were separated into 4 rows (3 photos per row) and were identified by a reference number printed above them (see Fig. 1). The 24 photos were in random order.

Each of the eight A4 response sheets had a question printed at the top of the page which made reference to the target property (e.g.: How much does this object show the property *large*?). It was explained that the task was to rate the extent to which the target properties were present, using a 7-point scale for each object. A score of 6 indicated that the property was maximally evident and 0 that it was not manifested at all.

The order of the response sheets was randomized between participants, but in any case we ensured that sheets referring to two contrary properties (e.g. *wide* and *narrow*) did not immediately follow each other. Participants were free to start from whichever object they wanted and to follow the order they preferred. It was made clear that the images were to be used only as 'evokers' of the object and that they were not being asked to judge the size of the object in the photo but the size of the real life object.

4.1.3. Stimuli

The objects were everyday objects and the aim was to cover a broad range of sizes. The set of objects selected were: a car, a pair of binoculars, an oak tree, a pair of scissors, a die, a handbag, a motorbike, a park bench, a pair of sunglasses, a cork, a pipe, an umbrella, a phone, a slide, a pencil sharpener, a chair, a stone archway, a rugby ball, a box of matches, a supermarket trolley, a tennis shoe, a laptop, a washbasin, the nave of a church (Fig. 1).

As stated in the note on methodology in the introduction to the study, one of the parameters estimated by Rasch models (namely, the item separation index) allowed us to control the distribution of the stimuli along the variable studied and therefore to test whether the stimuli used were a representative sample expressing different phenomenal sizes.

Fig. 1. The set of stimuli used in Study 1.

4.2. Results

We first explored the dimensionality of the data using an exploratory factor analysis. Since the distribution of the responses (tested by means of Kolmogorov–Smirnov tests) was significantly different from a normal distribution for most of the stimuli in the 8 properties, Principal Component Analysis (PCA) was used. For each pair of opposite properties, the ratings of one of the two properties were reversed according to the typical procedure used for testing whether reversed items and straightforward items fall in the same factor. Given that we had a 7 point scale (the percentage of use of the 7 points is reported in Appendix 2), for one of the two poles, e.g. *low*, we did 7 minus the rating of *low* given by participant and thus obtained a measurement which was 'reversed', i.e. expressed in terms of *high*, and comparable with the rating actually given in terms of *high*. Four matrices of data were thus obtained, each consisting of both 'straightforward ratings' (e.g. *high*) and the reversed rating for the opposite pole (e.g. the calculated reversed rating for *low*). If the two series of ratings refer to the same characteristic, they should jointly contribute to the same components and the structure of the data should show that the two series work together and are not distinct. Conversely, if the two ratings do in effect presuppose different underlying components, we should find that they correlate separately with *different* components.

For each pair of properties, the scree plots (Cattel, 1966) revealed that five or six Principal Components (PCs) were needed — they accounted for a percentage of total variance in between 53% and 69%, depending on the pair (Table 2). An analysis of the component loadings, i.e. the correlation coefficients between the variables and the PCs, revealed that after Varimax rotation the two series of judgments load separately in most of the PCs, including those which account for most of the variance (i.e. 1st and 2nd PCs). As shown in Table 2, the

Table 2
Percentage of total variance accounted for by the first 5–6 PCs extracted, for each of the matrices studied in Study 1 (each matrix was made up of 48 descriptions per subject, i.e. the description of the 24 stimuli in terms of the "straightforward" property and the reversed judgments for the opposite property). For each PC the number of items (stimuli_judgments) contributing to the identity of the PC with loadings bigger than |0.40| is also reported. The items referring to the two series of responses have been kept distinct in order to make evident which PCs are unbalanced (or even exclusive) in terms of one but not the other series of judgments.

	1st PC	2nd PC	3rd PC	4th PC	5th PC	6th PC	Total variance
% of total variance (*high–low*)	26.12	19.31	9.66	6.31	4.83	2.94	69.00
Stimuli judged for "high" with loadings > |0.40|	19	0	0	7	4	0	
Stimuli judged for "low" with loadings > |0.40|	0	16	9	0	0	3	
% of total variance (*long–short*)	20.44	16.10	9.25	5.29	4.81	3.08	58.94
Stimuli judged for "long" with loadings > |0.40|	0	17	0	6	4	0	
Stimuli judged for "short" with loadings > |0.40|	17	0	5	0	0	1	
% of total variance (*wide–narrow*)	22.03	14.81	7.22	5.26	3.75	–	53.12
Stimuli judged for "wide" with loadings > |0.40|	18	0	6	0	2	–	
Stimuli judged for "narrow" with loadings > |0.40|	0	15	0	4	1	–	
% of total variance (*large–small*)	23.93	15.97	11.19	6.66	3.54	–	61.31
Stimuli judged for "large" with loadings > |0.40|	14	0	0	6	2	–	
Stimuli judged for "small" with loadings > |0.40|	0	12	7	0	0	–	

loadings>|0.40| in the first PC referred *exclusively* to one of two series of judgments (e.g. in the pair *high–low*, out of the 19 variables with loadings>0.40, all refer to judgments of *high* and none to judgments of inverted *low*). Conversely, loadings>|0.40| in the second PC always referred to the opposite series of judgments (e.g. in the pair *high–low*, out of the 16 variables with loadings>0.40, all refer to reversed judgments of *low*). This aspect of variables with a high correlation in a PC exclusively referring to one of the two series of judgments was also found for PCs with smaller eigenvalues (i.e. 3rd and 4th PC).

The results of this exploratory analysis therefore suggest that various latent constructs lay under the two series of responses (the "straightforward" ratings and the reversed opposite ratings), and this means that the structure of the data is in any case not simple. Furthermore, these various latent constructs correlated separately with one or the other series and this is not compatible with the idea of a common continuum — or many common continua — underlying the two series of ratings.

Confirmatory Factor Analyses (CFAs) were carried out to test the null hypothesis that the fit of the one-factor model is not worse than that of the bi-factor model (the analyses were performed with Mplus version 6.11). Two confirmatory factor analyses (CFAs) were carried out to test the validity of *i*) a one-factor model that simply loaded all relevant items (i.e. the items with loadings>|0.50|) onto a single latent construct; *ii*) a two-correlated-factor model that loaded all relevant items referring to a property (e.g. *high*) onto one factor and the relevant items referring to the opposite property (e.g. reversed-*low*) onto the other factor. Since the Chi-square statistic is dependent on sample size, two relative fit indices were considered instead: the Non-normed Bentler–Bonnett fit index (TLI) and the Comparative Bentler fit index (CFI), as they both perform well with small and large samples. Values above 0.95 are usually considered satisfactory (Schermelleh-Engel, Moosbrugger, & Muller, 2003). The fit of the two-factor model was compared with that of the one-factor model to select the best-fitting model, using both a qualitative evaluation of the fit indices and the Akaike information criterion (AIC). As suggested by Schermelleh-Engel et al. (2003), given a set of models for the same data, AIC can be used to compare the competing models: the model with the minimum AIC value is regarded as the best-fitting model.

Since the PCA had revealed that at least 6 factors were needed to explain a percentage of variance between 53% and 69%, we did not expect the two-factor model to have a satisfactory fit. However, we expected it to have a better fit than the one-factor model. This was the case for all the properties. Although the Chi square was significant for both models ($p<0.05$), the TLI and CFI were always higher for the two-factor model than the one-factor model (and AIC is smaller in the first model than the second) — see Table 3. Both indices are however far from 0.95 (acceptable fit). A six-factor model was then tested, based on the results of the PCA (as for the previous models tested, we loaded on each factor the relevant items, i.e. those with loadings>|0.50|). For three out of the four pairs of opposites analyzed the 6-factor model (see third column of Table 3) had a satisfactory fit (TLI and CFI>0.95; AIC is much smaller for this than the other two models).

To further test the dimensionality, we applied Andrich Extended Rating Scale Model (Andrich Extended RSM), which is the extension of Rasch Dichotomous Models to rating scales (Andrich, 1988; Andrich & van Schoubroeck, 1989; Andrich & Luo, 1993). As mentioned in the Introduction, this analysis allowed us to make some additional verifications. The analyses were performed using RUMM 2020, developed by Andrich, Sheridan and Luo (2005)[2].

Table 3
Indices of fit associated with the three CFAs conducted in Study 1 (testing the one-factor, two-factor and six-factor models — for a description of the models see the main text of the paper).

Properties	Fit indices	CFA (1factor)	CFA (2factors)	CFA (6factors)
High–low	Non-normed Bentler–Bonnett fit index (TLI)	0.68	0.69	1.13[a]
	Comparative Bentler fit index (CFI)	0.72	0.70	1.00[a]
	Akaike (AIC)	5.37	12.82	1.056[a]
Wide–narrow	Non-normed Bentler–Bonnett fit index (TLI)	0.76	0.71	1.23
	Comparative Bentler fit index (CFI)	0.79	0.73	1.00
	Akaike (AIC)	2.86	7.78	1.02
Long–short	Non-normed Bentler–Bonnett fit index (TLI)	0.73	0.69	1.19
	Comparative Bentler fit index (CFI)	0.77	0.71	1.00
	Akaike (AIC)	3.3	8.57	1.19
Large–small	Non-normed Bentler–Bonnett fit index (TLI)	0.83	0.67	1.22
	Comparative Bentler fit index (CFI)	0.86	0.70	1.00
	Akaike (AIC)	1.92	10.1	1.22

[a] Note: only for high–low was the best fitting model (which these values refer to) the four-factor model.

I) *Stimuli (and participants) fall along continua and are well distributed along each continuum*. For each of the 8 target properties, it was verified that the Infit Mean Square[3] for each stimulus (and each participant) fell inside the critical range (0.75<Infit Mean Square<1.33). This meant that the model-data fit was guaranteed and therefore that stimuli and participants could be scaled along a linear continuum, one for each of the 8 target properties[4].
We further tested how well the stimuli (and participants) were distributed along each of the 8 continua by means of reliability coefficients. Testing whether the set of stimuli (and the sample of participants) was homogeneously distributed along the continuum was relevant in order to exclude the event that the lack of unidimensionality between the two opposite scalings (eventually found in the next step of the analysis) was due to the idiosyncratic behavior of some of the stimuli (or participants). The Item and Person separation reliability coefficients (which vary from 0 to 1) were excellent for all of the target properties: the Item separation index ranged from 0.85, for *large*, to 0.91, for *small* and *high*; the Person separation index ranged from 0.86, for *large*, to 0.95, for *low*. This confirmed that there was no anomalous behavior for judgments of some of the items or of some of the participants.
The goodness of these two indices for all the 8 properties is a theoretically important result. It has often been emphasized by linguists, as a consequence of observing the linguistic usage of gradable adjectives, that they are always comparative in nature, although not necessarily in form, and therefore sensitive to the reference point contextually construed (Cruse, 1986; Paradis, 2008). It has also been emphasized

[2] In conducting Rasch analyses, RUMM2020 uses the "Pairwise Conditional Estimation procedure" which generalizes the equation for one pair of items, in which the person parameter is eliminated, to all pairs of items taken simultaneously. Pairwise estimation is a conditional estimation in the sense that the person parameters are eliminated while the item parameters are estimated (see Zwinderman, 1995).

[3] Rasch himself suggested the use of chi-square fit statistics to determine how well any set of empirical data met the requirements of his model. Rasch analysis programs usually report fit statistics as two chi-square ratios: Infit and Outfit Mean Square statistics. Infit is an information-weighted sum (Wright, 1984; Wright & Masters, 1981).

[4] The risk put forward by Wood (1978) that Rasch Models might provide an indication of good fit even when items measuring independent characteristics were in fact tested as lying on the same continuum, does not apply to our data. The "items" used in our studies were homogeneous (i.e. all photographic representations of everyday objects). Participants were not asked to provide a judgment for items which might be different in nature and require different abilities, but only to rate the degree to which a certain spatial property (always the same for all the items) was present. The condition of potential error due to the independence of the items presented can thus be excluded in this case and the resulting good fit can be considered a real indication that the structure of the data manifests a continuous latent construct, as identified by the model.

that the nature of the scale itself is subject to variable contextual construals (Cruse & Togia, 1995); an adjective such as 'short' may in some contexts be construed as operating over the same scale as 'long' (e.g. 'a short holiday/a long holiday') but in other contexts may be construed as operating over an independent scale oriented in the opposite direction — e.g. 'How short is it?'). These are certainly interesting observations concerning the linguistic behavior of adjectives and can help us to understand which cognitive processes of re-centering take place when different expressions are used. However, our results show that an invariant and coherent perceptual scale underlying participants' judgments of sizes is present. The ratings given by participants concerning the degree to which a property is visible in 24 different objects was not as arbitrary and context dependent as one might expect based on linguistic considerations. Ratings lay on a continuous latent construct, that is to say that invariances between the sizes of different objects were recognized. In other words, if half of the story is that we can say, describing an ant walking on our desk, that 'it is enormous' (as compared to, say, a tiny ant), the other half of the story is that we however recognize that it is in any case a very small object: the motor plan that we would activate for grasping it is that typical for grasping entities of a few millimeters (an ant, a crumb, a minuscule seashell or a ladybird) and we would never use it for an umbrella to protect us from the rain.

II) *The unidimensionality/non-unidimensionality of opposite scales*. In order to verify whether judgments for the two opposite properties lay on the same latent continuum, we focused on the δ_i values obtained from the Andrich Extended RSM (that is on the scalings of the set of stimuli according to the degree to which they manifest the target property). Since we were working with opposite properties, we did the same as for the factor analysis, i.e. we analyzed "straightforward" judgments (raw data) for one property and came up with the δ_i logit for that property (e.g. *large*). We then "reversed" judgments for the opposite property (by doing 7 minus raw rating) and came up with the δ_i logit (e.g. reversed-*small*). The two series of δ_i values were then compared by applying a procedure suggested by Wright and Stone (1979, pp. 94–95) and Bond and Fox (2001) in order to determine whether two scalings comply with an identity function (within the accepted 95% confidence band). According to this procedure, the two scales are first equated to ensure that they have the same average (in our case this was not necessary since the δ_is have a common average, i.e. 0) and then plotted (as shown in Fig. 2). If the two scales lie on the same continuum, the slope of the linear function representing the estimates should be around 45°. As shown in Fig. 2, no identity relationship was found for any of the pairs of properties examined — i.e. most of the plotted estimates fell outside the confidence band (the hatched lines).

Therefore, the results of Factor analysis and Andrich Extended RSM consistently proved that the two series of ratings are not measurements of the same dimension. A negative correlation between the two series of opposite ratings, as predictable, was found (mean correlation between *long* and *short*: $r = -0.61$; between *wide* and *narrow*: $r = -0.77$; *large* and *small*: $r = -0.78$; and *high* and *low*: $r = -0.75$). However, both CFA and Andrich Extended RSM demonstrated that these correlations were

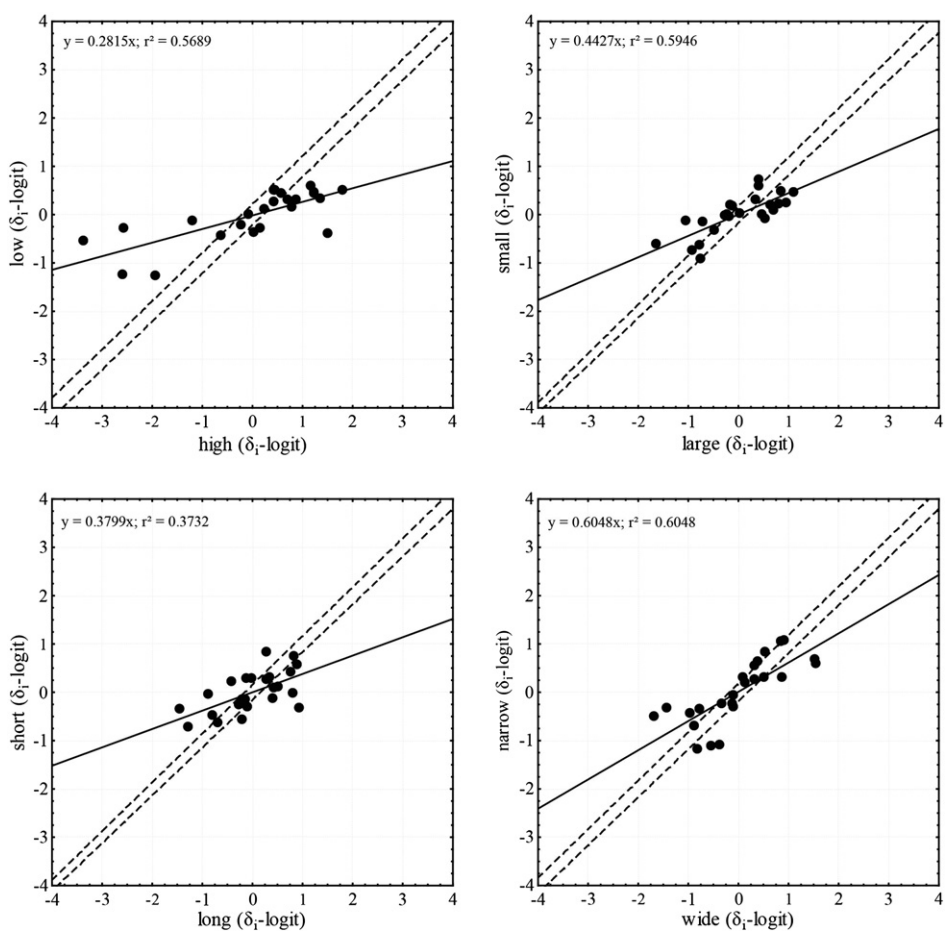

Fig. 2. The relationship between *high/low*, *large/small*, *long/short* and *wide/narrow* (in δ_i logit), according to the results of Study 1. In each graph the area between the broken lines represents the area where the hypothesis that the two scales measure the same characteristics is accepted (95% confidence band).

not strong enough to consider, in practice, the two opposite judgments to be redundant and overlapping.

5. Study 2

The results from Study 1 suggest that *high*/*low*, *large*/*small*, *wide*/*narrow*, and *long*/*short* do not lie, respectively, on the same continuum. Since ratings were given by looking at pictures of objects, and not at the objects themselves, one might wonder whether these findings may simply be due to the method used. In Study 2, we sought a validation of the findings from Study 1 by means of a task involving the observation of real-life objects.

5.1. Method

5.1.1. Participants

112 undergraduate students at the University of Macerata (89 females and 23 males, mean age 21.4).

5.1.2. Procedure

The task was the same as in Study 1, except that participants were now asked to express their judgments when directly observing 24 objects. The objects were observed in a natural context (e.g. a car in the street, a park bench and oak tree in a park, a tennis shoe on a person's foot, a pair of scissors, a telephone and a laptop on an office desk, etc.). They described the objects one at a time, using all 8 target properties. The order of the properties was controlled in order to avoid two opposite properties immediately following each other.

5.1.3. Stimuli

Twenty-four real-life objects (the same as those presented in pictures in Study 1).

5.2. Results

As in Study 1, the unidimensionality of the 4 pairs of judgments was first explored by means of PCA (after Kolmogorov–Smirnov tests demonstrated the lack of normality of the distribution of responses for most stimuli) and then tested with CFA. For each pair of spatial opposites, the ratings of one of the two properties was reversed. Four matrices of data were thus obtained, each consisting of ratings given to both the "straightforward items" (e.g. *high*) and of the reversed judgments for the opposite pole (e.g. the reversed values calculated for *low*). The scree plots revealed that five or six PCs contributed to explain a percentage of total variance between 53% and 62%, depending on the pair (Table 4).

As shown in Table 4, the analysis of the component loadings revealed that the two series of judgments tend to load separately on the PCs which account for most of the variance (i.e. 1st and 2nd PC). This separation was also found for other PCs with smaller eigenvalues (i.e. 3 rd PC for *long–short*, *wide–narrow* and *large–small*, and 5th PC for *high–low* and *long–short*).

As in Study 1, three CFAs were carried out to test the validity of *i*) a one-factor model that simply loaded all relevant items onto a single latent construct, *ii*) a two-correlated-factor model that loaded all relevant items referring to a property (e.g. *high*) onto one factor and the relevant items referring to the opposite property (e.g. reversed-*low*) onto the other factor and *iii*) a six-factor model defined on the basis of the PCA. Although the Chi square was significant, the other indices (reported in Table 5) showed that the two-factor model provided a better fit than the one-factor model (CFI and TLI were always higher for the two-factor model than the one-factor model and AIC was smaller for the first than the second). As in Study 1, both CFI and TLI were however far from 0.95 (acceptable fit). The six-factor model identified based on the results of the PCA pointed to a satisfactory fit (TLI and CFI were above 0.97, AIC was much smaller than the other two models).

Table 4
Percentage of total variance accounted for by the first 5–6 PCs extracted, for each of the matrices in Study 2 (each matrix was made up of 48 descriptions per subject, i.e. the description of the 24 stimuli in terms of the "straightforward" property and the reversed judgments for the opposite property). For each PC the number of items (stimuli_judgments) contributing to the identity of the PC with loadings bigger than |.40| is also reported. The items referring to the two series of responses have been kept distinct in order to make evident which PCs are unbalanced (or even exclusive) in terms of one but not the other series of judgments.

	1st PC	2nd PC	3rd PC	4th PC	5th PC	6th PC	Total variance
% of total variance (*high–low*)	21.23	10.95	9.53	7.49	5.78	5.22	60.22
Stimuli judged for "high" with loadings>|0.40|	10	1	1	2	7	1	
Judgments of "low" with loadings>|0.40|	2	12	2	1	0	1	
% of total variance (*long–short*)	24.73	11.13	8.44	6.56	6.03	4.66	61.59
Stimuli judged for "long" with loadings>|0.40|	3	0	7	6	0	1	
Stimuli judged for "short" with loadings>|0.40|	0	12	0	5	5	2	
% of total variance (*wide–narrow*)	21.40	11.22	8.47	6.47	5.51	–	53.09
Stimuli judged for "wide" with loadings>|0.40|	0	8	6	2	6	–	
Stimuli judged for "narrow" with loadings>|0.40|	10	1	0	0	1	–	
% of total variance (*large–small*)	24.34	11.87	8.96	6.56	4.88	–	56.63
Stimuli judged for "large" with loadings>|0.40|	0	8	5	2	6	–	
Stimuli judged for "small" with loadings>|0.40|	14	2	0	1	2	–	

Therefore, as in Study 1, the results of the factor analysis did not support the hypothesis that responses in terms of opposite properties manifest a simple unidimensional structure. To further test the dimensionality of the responses we applied the Andrich Extended RSM.

I) *Stimuli* (*and participants*) *fall in continua and are well distributed along each continuum*. As in Study 1, it was verified that the Infit Mean Square for each stimulus (and each participant) fell inside the critical range (0.75<Infit Mean Square<1.33). The model-

Table 5
Indices of fit associated with the three CFAs conducted in Study 2 (testing the one-factor, two-factor and six-factor models — for a description of the models see the main text of the paper).

Properties	Fit indices	CFA (1factor)	CFA (2factors)	CFA (6factors)
High–low	Non-normed Bentler–Bonnett fit index (TLI)	0.64	0.71	1.19
	Comparative Bentler fit index (CFI)	0.68	0.73	1.00
	Akaike (AIC)	4.53	3.26	1.62
Wide–narrow	Non-normed Bentler–Bonnett fit index (TLI)	0.78	0.67	1.13
	Comparative Bentler fit index (CFI)	0.73	0.71	1.00
	Akaike (AIC)	3.11	5.19	1.83
Long–short	Non-normed Bentler–Bonnett fit index (TLI)	0.72	0.70	1.15
	Comparative Bentler fit index (CFI)	0.79	0.77	1.00
	Akaike (AIC)	4.4	5.1	1.27
Large–small	Non-normed Bentler–Bonnett fit index (TLI)	0.69	0.81	1.26
	Comparative Bentler fit index (CFI)	0.74	0.80	1.00
	Akaike (AIC)	2.97	2.71	1.83

data fit thus ensures that stimuli and participants, for each of the 8 properties, were well distributed along a continuum. Both the Item and the Person separation reliability coefficients were excellent for all of the 8 target properties: the Item separation index ranged from 0.88, for *wide*, to 0.93, for *small* and *narrow*; the Person separation index ranged from 0.91, for *wide*, to 0.94, for *small* and *long*. This, as in Study 1, gives support to the idea that participants' judgments of *high*, *low*, *large*, *small*, *wide*, *narrow*, *long* and *short* each followed an ordinal additive structure (i.e. a continuum) which was invariant across the stimuli presented. This led us to ask whether the continua pertaining to two opposite properties are in effect a single common continuum.

II) *The unidimensionality/non-unidimensionality of the opposite scales.* Replicating the procedure used in Study 1, we worked on straightforward ratings for one property and reversed ratings for the other, estimating the corresponding δ_i values. If the two interval scales (e.g. the scale for *high* and the reversed scale for *low*) refer to the same latent construct, the slope of the linear function representing the scalings of the two properties (e.g. *high* and reversed *low*) should be around 45°. As shown in Fig. 3, in no cases did the estimates referring to one pole lie on the same continuum as the estimates referring to the opposite pole.

Thus, even in ecological conditions, the factor analyses and Andrich Extended RSM confirm that judgments of *high*/*low*, *large*/*small*, *wide*/*narrow* and *long*/*short* do not conform to the hypothesis of a latent common dimension.

6. Study 3

A third study was conducted in order to understand whether unidimensionality (absent in Studies 1 and 2) would be present if, instead of using discrete objects (a die, a pair of scissors, a motorbike or a wash-basin…), only 'extensions' in space were considered.

6.1. Method

6.1.1. Participants

120 undergraduate students at the University of Verona (81 females and 39 males, mean age 22.3) participated in the study.

6.1.2. Procedure

The experiment was conducted in the open air. Two small planks of wood painted red (30 × 10 cm) were used as markers on the ground. One of the two planks (plank 1) remained in the same position, while the other was moved by the experimenter into 24 different positions.

In condition A, plank 1 was placed horizontally in front of the observer and the second plank (plank 2) was moved into 24 different positions along the sagittal axis (see Fig. 4, condition A). Participants were asked to describe how *long* (or *short*), the space between the two planks was, using a 7 point scale. In condition B, plank 1 was placed perpendicular to the observer, 20 cm to the left. Plank 2 was moved into 24 different positions along the coronal axis to the right (see Fig. 4, condition B). Participants were asked to evaluate how *wide* (or *narrow)* the space was, again using a 7 point scale.

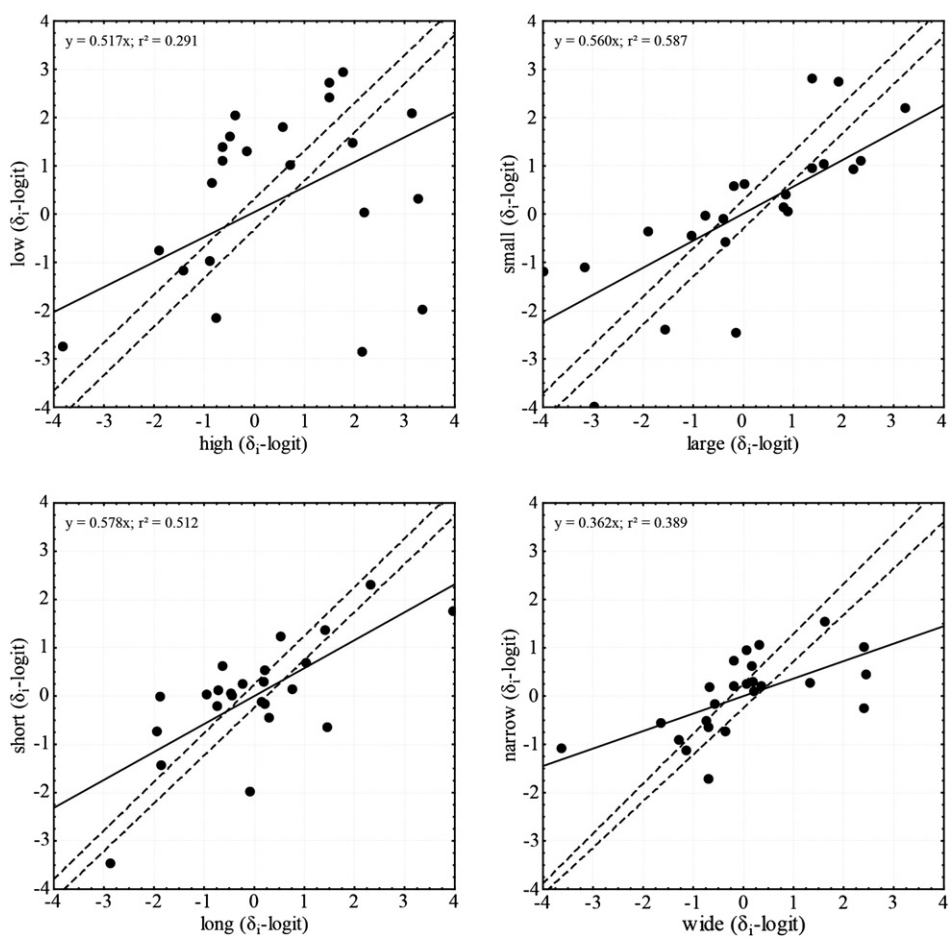

Fig. 3. The relationship between *high*/*low*, *large*/*small*, *long*/*short* and *wide*/*narrow* (in δ_i logit), according to the results of Study 2. In each graph the area between the broken lines represents the area where the hypothesis that the two scales measure the same characteristics is accepted (95% confidence band).

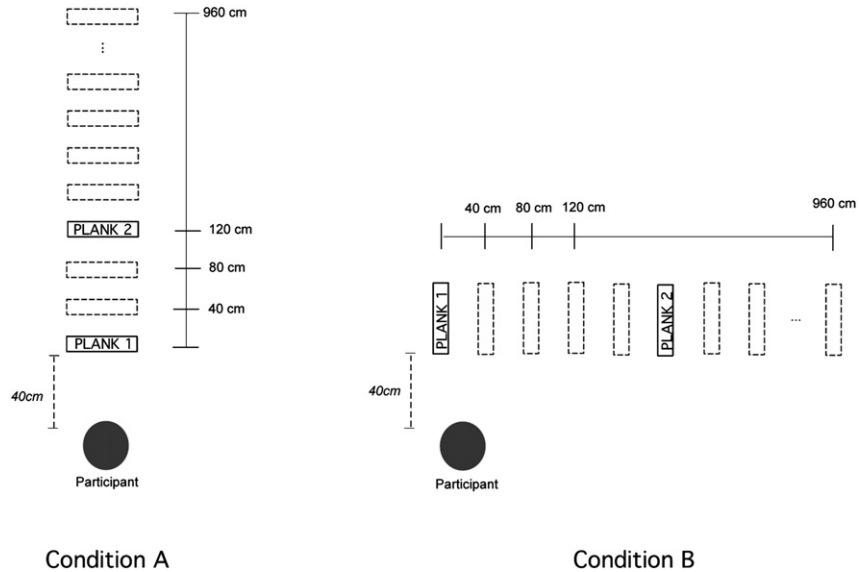

Fig. 4. Conditions used in experiment 3 to study the properties *long/short* (condition A) and *wide/narrow* (condition B). Plank 1 remained in the same position while the plank 2 was moved into 24 different positions (here represented by the rectangles in broken lines) with respect to the participant.

The order of the properties to be judged (*long*, *short*, *wide*, and *narrow*) varied between participants, who described the whole set of 24 stimuli (presented in random order) for each target property before moving on a new target property. As in the previous two studies, participants were never asked to rate two opposite properties one immediately after the other.

6.1.3. Stimuli

Twenty-four extensions (ranging from 40 cm to 960 cm, at intervals of 40 cm) for each of the 2 dimensions studied: *long/short* (Fig. 4, condition A) and *wide/narrow* (Fig. 4, condition B).

6.2. Results

As in the previous two studies, we first applied PCA to the data in order to explore whether the structure of the latent constructs underlying the raw ratings would indicate that the two series of opposite judgments loaded separately onto different PCs. The scree plots revealed that four PCs were required: they account for 73% of total variance for l*ong–short*, and 69% for *wide–narrow*. In this case too, as shown in Table 6, one series of ratings (with only one exception for the first PC regarding *wide–narrow*) loaded exclusively on the first PC while the other loaded on the second PC. The same was found for the 3rd PC for *long–short* and the 4th PC for *wide–narrow*.

As in the previous studies, three CFAs were carried out to test the validity of *i*) a one-factor model that simply loaded all relevant items onto a single latent construct, *ii*) a two-correlated-factor model that loaded all relevant items referring to a property (e.g. *high*) onto one factor and the relevant items referring to the opposite property (e.g. reversed-*low*) onto the other factor and *iii*) a four-factor model defined on the basis of the PCA. The Chi square was significant for all three models, but the other indices (shown in Table 7) indicated that the two-factor model provided a better fit than the one-factor model (CFI and TLI were always higher for the 2-factor model than the 1-factor model and AIC was smaller in the first than the second). However both CFI and TLI were far from 0.95 (acceptable fit). The four-factor model identified based on the results of the PCA pointed to a satisfactory fit (TLI and CFI were above 0.97, AIC was much smaller than the other two models).

As in the previous two studies, to further test the dimensionality of the responses we applied the Andrich Extended RSM.

I) *Stimuli* (and p*articipants*) *fall along continua and are well distributed along each continuum*. The Infit Mean Square for each stimulus (and each participant) fell inside the critical range (0.75<Infit Mean Square<1.33). Therefore, the stimuli (and participants) fell along continua for each of the 4 properties analyzed. The Item and the Person separation reliability coefficients were excellent

Table 6
Percentage of total variance accounted for by the first 4 PC extracted, for each of the two matrices in Study 3 (each matrix consisted of 48 descriptions per subject, i.e. the description of the 24 stimuli in terms of the "straightforward" property and the reversed judgments for the opposite property). For each PC the number of items (stimuli_judgments) contributing to the identity of the PC with loadings bigger than |.40| is also reported. The items referring to the two series of responses have been kept distinct in order to show which PCs have an unbalanced (or even exclusive) characterization in terms of one but not the other series of judgments.

	1st PC	2nd PC	3rd PC	4th PC	Total variance
% of total variance (*long–short*)	38.86	19.84	8.18	6.02	72.96
Judgments of "long" with loadings> |0.40|	0	18	10	0	
Judgments of "short" with loadings> |0.40|	18	0	0	1	
% of total variance (*wide–narrow*)	30.07	25.99	7.92	4.83	68.82
Judgments of "wide" with loadings> |0.40|	18	0	1	0	
Judgments of "narrow" with loadings>|0.40|	1	19	1	5	

Table 7
Indices of fit associated with the three CFAs conducted in Study 3 (testing the one-factor, two-factor and six-factor models — for a description of the models see the main text of the paper).

Properties	Fit indices	CFA (1factor)	CFA (2factors)	CFA (4factors)
Wide–narrow	Non-normed Bentler–Bonnett fit index (TLI)	0.65	0.64	1.19
	Comparative Bentler fit index (CFI)	0.71	0.70	1.00
	Akaike (AIC)	5.13	4.97	1.62
Long–short	Non-normed Bentler–Bonnett fit index (TLI)	0.78	0.79	1.02
	Comparative Bentler fit index (CFI)	0.79	0.81	1.00
	Akaike (AIC)	4.07	3.95	1.26

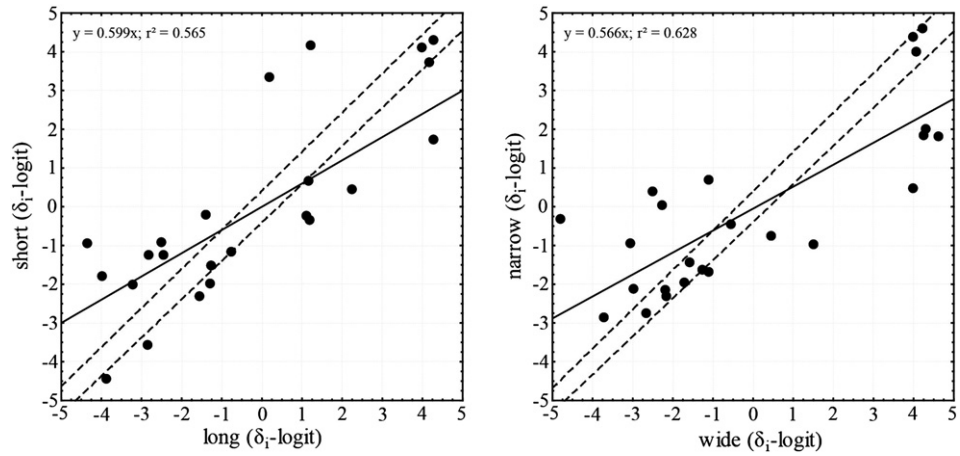

Fig. 5. Relationship between *long*/*short*, *wide*/*narrow* (in δ_i logit). The broken lines represent the area where the hypothesis that the two scales measure the same characteristics is accepted (95% confidence band).

for all of the 4 target properties (the Item separation index ranged from 0.95, for *wide* and *narrow*, to 0.97, for *long*; the Person separation index ranged from 0.95, for *wide*, to 0.99, for *long)*, confirming a good positioning of the stimuli (and the subjects) all along each continuum.

II) *The unidimensionality*/*non-unidimensionality of the opposite scales*. As shown in Fig. 5, when comparing the scalings of stimuli referring to one series of estimates (e.g. δ_i values for *long*) and its reversed opposite (reversed δ_i values for *short*), the two series turned out not to lie on the same continuum also in the case of this condition. The same was found when comparing the two scalings for *wide* and (reversed) *narrow*.

Therefore, when simple extensions rather than objects were considered, the results of both factor analyses and Andrich Extended RSM were consistent in suggesting that the two series of ratings were not unidimensional.

7. Joint analysis

The studies consistently proved that *high*/*low*, *large*/*small*, *wide*/*narrow*, and *long*/*short* cannot be considered as inverse gradations of the same characteristic. How can these findings be explained? A linguist might wonder if this result simply indicates that it is difficult to apply the marked adjective to the whole set of object sizes considered. By definition, this would not be the case with the unmarked adjective, considered "unmarked" since it can cover the whole dimension. As a consequence, participants may have found it easy to describe the whole set of objects in terms of different gradations of, say, *high*, but more difficult to describe the whole set of objects in terms of gradations of *low*. This hypothesis is not however compatible with our results. Signs that the whole range of sizes was difficult to judge in terms of the marked pole (*short*, *small*, *narrow* and *low)* would have been evident in the Andrich Extended RSM if the values of Infit had been outside the critical range and the values of the Item and Person reliability separation coefficients had not been excellent for the scales related to the marked properties. In contrast, these indices were excellent for all the properties studied in all the studies. Markedness does not therefore provide the key to our results.

We suggest that the reason for our findings is the implicit association of the properties studied with other properties. In perception, the interaction between properties is the rule. Geometrical optical illusions are extreme examples of a broader phenomenon that represents the non-isolated behavior of qualities. This characteristic of non-independence is recognized also when modeling conceptual spaces: "the dimensions of a conceptual space should not be seen as totally independent entities, rather they are correlated in various ways since the properties of the objects modeled in the space co-vary. For example, in the domain of fruits the ripeness and the color dimensions co-vary" (Gärdenfors, 2007, p. 3).

Thus, when participants estimated to what extent a given object appeared to be *high* or *low*, or *large* or *small*, they were not measuring the actual physical size of objects but describing the way in which they appeared to them and other dimensions might therefore influence their judgments. These implicit associations do not necessarily affect the two poles in a corresponding way (e.g. *wide* might interfere more with judgments of *low* than with *high* or might interfere only with one and not the other property) and this may lead to a lack of correspondence between the two scalings of opposite properties.

At a first glance, this hypothesis seems to be confirmed when the order of objects within the scaling of the 24 stimuli (based on δ_i) is considered. For instance, as shown in Fig. 6, where the labels of some items have been added to the panel of Fig. 3 regarding *high*–

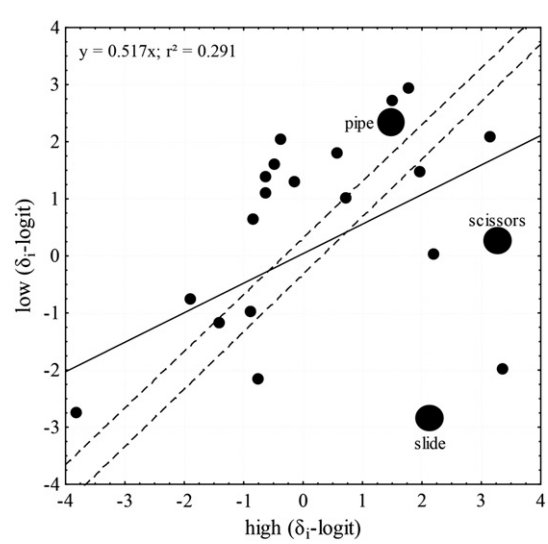

Fig. 6. Position of the stimuli slide, scissors and pipe with respect to the δ_i scalings of *high* and inverted-*low* in Study 2 (the figure is a modification of Fig. 3). With respect to the 0 point which represents average difficulty in displaying the property, negative values of δ_i logit characterize the stimuli that clearly display the target property, positive values of δ_i logit characterize the stimuli where it is difficult to perceive. For the comments, see text.

low, in terms of difficulty in perceiving the property *high*, the scissors came top of the list, followed the slide and the pipe. These objects have in fact high positive δ_i logit in the scale of *high* (i.e. it was difficult to see *high* when looking at these objects). It is of course evident that a slide is physically much higher than a pipe. However, our results demonstrate that it was very difficult to see this property in the slide — as much as (and even a bit more than) in the pipe. For the inverse ratings of *low*, the pipe and the scissors were at one end of the scale — they had positive values of inverted-*low*, which means that they received high ratings of *low* (or that it was easy to see *low* when looking at them) — while the slide was at the opposite side of the scale which means that it received low ratings of *low* (i.e. it was difficult to see low when looking at it).

A full test of the interactions between the 8 properties studied with any other spatial properties would require a correlation matrix that goes well beyond that provided by our studies (we should in fact also be able to study the interaction between perceptions of shape and size). However, we can start by exploring the correlations between the 8 series of judgments. In order to do this, two PCAs were conducted, on Study 1 and Study 2 respectively. We focused on these two studies which consider 4 pairs of properties rather than only two. The analysis was applied to the δ_i values derived from the application of the Andrich Extended RSM. The advantage of working on δ_i values rather than on raw data is that δ_i values are logit scales, therefore there are no problems related to comparing scales with different shapes or different units of measure. By analyzing δ_i values we can focus on variability between stimuli, clear of any variability related to participants (which is expressed by the β_n values).

Two factors account for a considerable amount of the total variance in both studies (80.63% in Study 1 and 77.35% in Study 2). The eigenvalues of the two factors are very similar between studies, with the first factor explaining the high percentage of total variance (71% in Study 1 and 63% in Study 2) and the second factor explaining the 10% of variance in Study 1 and the 14% in Study 2.

As shown by the scatter plots representing the 2 PCs (Fig. 7), the first PC refers to the overall spatial extension of the objects. On the right we found all the properties referring to reduced extensions, i.e. *small*, *narrow*, *low* and *short* (with loadings ranging from 0.84 to 0.92 in Study 1 and from 0.76 to 0.90 in Study 2). On the left we found the properties referring to greater extensions: *large* and *wide* have the highest positive loadings (respectively, −0.91 and −0.92 in Study 1 and −0.89 and −0.74 in Study 2). These two variables load specifically on the first PC since they have loadings of less than |0.12| in the second factor. *High* and *long*, which have loadings ranging from −0.77 to −0.60 in the 1st PC, also have high loadings in the 2nd PC (*high*: 0.50 in Study 1 and −0.64 in Study 2; *long*: −0.63 in Study 1 and 0.70 in Study 2). Only these two variables load on the second PC.

The first and most important factor therefore has to do with the overall extension of an object. In line with this, the highest correlations within this PC were not specifically between canonical opposites, but similarly high negative correlations were found between the properties occupying the opposite side of the PC. For instance, in both Studies 1 and 2, no significant differences were found between the negative correlation of *low* with its canonical opposite *high* (r = −0.75 in Study 1 and r = −0.64 in Study 2) or its correlation with *large* (r = −0.73 in Study 1 and r = −.69 in Study 2) or *wide* (r = −0.73 in Study 1 and r = −0.77 in Study 2). Similarly, the negative correlation of *large* with its canonical opposite *small* (r = −78 in Study 1 and r = − 77 in Study 2) was not significantly higher than its correlation with *short* (r = −0.84) or *low* (r = −.73) in Study 1 and with *short* (r = −74 in Study 2), *low* (r = −70) and *narrow* (r = −77) in Study 2. The negative correlation of *short* with *large* (r = − 74) in Study 2 was as high as with its canonical opposite *long* (r = −72) and even significantly higher in Study 1 (*short-long*, r = −0.61; short-*large*, r = −84; the difference was significant at $p<0.05$).

The second factor concerns the orientation of the object, as expressed by its elongation axis. As the 2nd PC suggests, judgments of *high* and *long*, which on the one side are related to a generic judgment of extension (loading on the 1st PC), also vary based on the elongation of the object and therefore its perceived spatial orientation along the vertical or horizontal plane. In this sense we do not necessarily need to be in front of an extended object to perceive it as *high* or to perceive it as *long*. Objects which are perceived as *long* have a main extension oriented close to or along the horizontal plane. Objects which are perceived as *high* have a main elongation along the vertical. According to the indications provided by the second PC, we would expect *vertical* and *horizontal* to correlate, respectively, with *high* and *long* more than with the other spatial properties in the first PC.

8. Final discussion

The purpose of the studies presented in this paper was to verify whether perceptual judgments of *high*/*low*, *large*/*small*, *wide*/*narrow* and *long*/*short* applied to ecological objects or to spatial extensions behave as inverse measurements of the same characteristics or whether they presuppose at least partially distinct continua. This research aimed to contribute to the investigation of the cognitive structure of opposites which has been addressed in part in the field of cognitive linguistics by means of observation of linguistic mechanisms in conjunction with potentially correlated cognitive constraints (for an overview, see

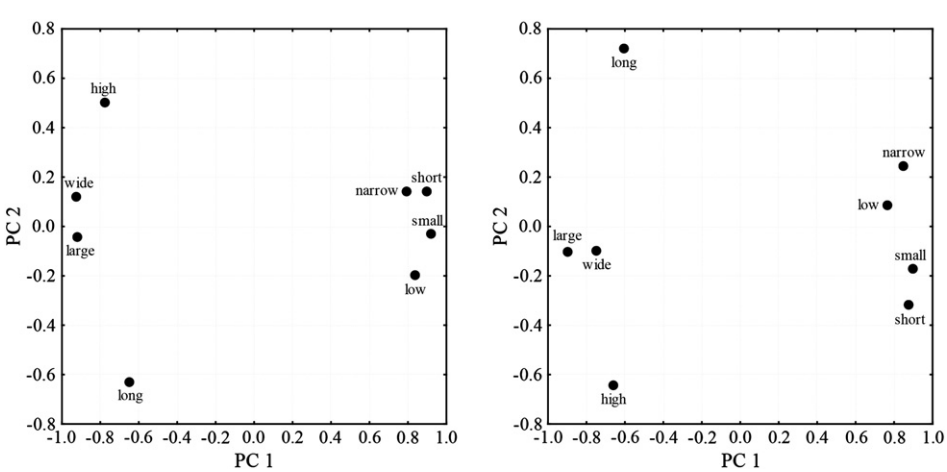

Fig. 7. Scatterplots of the PCA on the 8 variables in Study 1 (on the left) and 2 (on the right) where the first 2 PCs are plotted. The congruence coefficients proved that, although inverted in sign, the 2nd PCs measure the same in the two studies.

Muehleisen, 1997; Murphy, 2003). A recent approach to the study of opposition is founded on the use of perceptual and phenomenological tasks (Bianchi & Savardi, 2008a,b; Bianchi et al., 2011; Savardi, Bianchi, & Burro, 2009; see also the study of the perceptual dimensions of tactile textures carried out by Hollins et al., 1993 and by Picard et al., 2003).

The joint application of factor analysis and Andrich Extended RSM allowed us to investigate the structure of the dimension/dimensions underlying opposite series of ratings given for the same set of stimuli. The three studies consistently showed that the variance in data is not compatible with a single continuum underlying the two series. The PCAs conducted on the raw judgments of the two opposite properties (one dealt with as a response to a straightforward item, and the other as a response to a reversed item) demonstrated that 4–6 PCs needed to be extracted to account for a percentage of total variance ranging from around 53% to 73% in the three studies, and that in any case the PCs with higher eigenvalues (and sometimes also others with smaller eigenvalues) showed exclusive or almost exclusive loadings for one and not the other series of judgments. In all three studies the CFAs confirmed that the one-factor model had the worst fit. Application of the Andrich Extended RSM enabled us to refine the analysis of the data, focusing on the variability of responses concerning the stimuli (δ_i values) as distinct from the variability of responses concerning the participants (β_n values). This revealed that the scalings of the single properties are in fact compatible with continuous monotonic distributions. However when the two scalings corresponding to the two opposite properties (the straightforward continuum, e.g. *high*, and the reversed continuum, e.g. reversed *low*) are compared, they do not lay on a single common continuum.

Our findings are in line with the many results found in psychological test development showing that items keyed in the same direction as the construct to be measured (straightforward items, e.g., *reliable*) and in the opposite direction (reverse worded items, e.g., *unreliable*) do not lead necessarily to the same (reversed) results. This is usually explained in terms of method bias (the literature about biases due to negatively worded items is well reviewed in Barnette, 2000). We have suggested that this absence of a plain correspondence between ratings of opposite properties might be grounded on the cognitive nature of opposites, which are not simply mirror images of each other because each property is affected by interactions with other properties which do not necessarily act symmetrically on the two poles. With respect to the spatial properties analyzed, a joint analysis of the eight properties revealed, in both Studies 1 and 2, that judgments referring to a lack of extension are strictly correlated to each other and that they vary as if a single common latent construct is presupposed but the same is not symmetrically true for the opposite poles. *Large* and *wide* have high correlations and in effect behave in opposition to judgments of lack of extension, as expressions of a great amount of extension. But, aside from the global extension factor, judgments of *long* and *high* also depend on the perception of a main elongation axis, independently of the absolute size of the object, and therefore depend on perceived orientation.

There is evidence that contrary properties are not simply two poles of a single monoscalar system, and this has emerged in various areas of psychological research (psychometrics, cognitive linguistics and now experimental phenomenology). This should stimulate a search for structures capable of modeling the reciprocal behavior of opposites and lead us to consider to what extent unidimensional linear models of dimensions can be generalized within the Cognitive Sciences. This potentially impacts on models of conceptual spaces developed in contemporary cognitive science (e.g. Gärdenfors, 2007), where the shape of dimensions is in theory modeled on psychophysical data conveying cognitive structure.

The lack of unidimensionality may go toward explaining why opposites are so pervasive in human cognition (Kelso & Engstrom, 2006; Jones, 2002) and so primal (children aged 3–6 months are already sensitive to contrastive properties such as *up–down*, *tight–loose*, *inside–outside*; see Casasola, 2008; Casasola, Cohen, & Chiarello, 2003; Quinn, 2005). It proves that two different labels refer in effect to two different (and partially independent) dimensions. The fact that two opposites are non-reducible may be one of the keys to understanding why people, over time, demonstrate that they have not been able to get rid of bipolar structures (in favor of a more economical uni-polar system) in their natural modeling of their own perceptual, social, affective experience of the world.

Acknowledgments

We would like to thank Stéphane Vautier and two anonymous reviewers for their valuable comments on an earlier version of the manuscript.

Appendix 1 and Appendix 2. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.actpsy.2011.08.003.

References

Andersen, E. B. (1982). Latent trait models and ability parameter estimation. *Applied Psychological Measurement*, *6*(4), 445–461.
Andrich, D. (1978a). A binomial latent trait model for the study of Likert-style attitude questionnaires. *British Journal of Mathematical and Statistical Psychology*, *31*, 84–98.
Andrich, D. (1978b). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573.
Andrich, D. (1978c). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, *2*, 581–594.
Andrich, D. (1988). *Rasch models for measurement*. Beverly Hills, USA: Sage.
Andrich, D., & van Schoubroeck, L. (1989). Statistical reasoning in psychometric models and educational measurement. *Journal of Educational Measurement*, *26*(1), 7–26.
Andrich, D., & Luo, G. (1993). A hyperbolic cosine latent trait model for unfolding dichotomous single-stimulus responses. *Applied Psychological Measurement*, *17*, 253–276.
Andrich, D., Sheridan, B. S., & Luo, G. (2005). *Interpreting RUMM2020*. Perth, WE: RUUM Laboratory PTY LTD.
Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement*, *60*(3), 361–370.
Bianchi, I., & Savardi, U. (2008a). *The perception of contraries*. Roma, IT: Aracne.
Bianchi, I., & Savardi, U. (2008b). Contrariety as perceptual relationship. *Gestalt Theory*, *2*, 106–120.
Bianchi, I., Savardi, U., & Kubovy, M. (2011). Dimensions and their poles: A metric and topological theory of opposites. *Language and Cognitive Processes*. doi:10.1080/01690965.2010.520943.
Billet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, *7*(4), 608–622.
Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
Bonnefon, J. F., Eid, M., Vautier, S., & Jmel, S. (2008). A mixed Rasch model of dual-process conditional reasoning. *Quarterly Journal of Experimental Psychology*, *61*(5), 809–824.
Bozzi, P. (1989). *Experimental phenomenology*. Bologna, IT: il Mulino.
Burro, R. (2009). Measuring in experimental phenomenology and carrying out phenomenological psychophysics: The case of contrary properties. In U. Savardi (Ed.), *The perception and cognition of contraries* (pp. 249–274). Milano, IT: McGraw-Hill.
Burro, R., Sartori, R., & Vidotto, G. (2011). The method of constant stimuli with three rating categories and the use of Rasch models. *Quality and Quantity*, *45*(1), 43–58.
Casasola, M. (2008). The development of infants' spatial categories. *Current Directions in Psychological Science*, *17*(1), 21–25.
Casasola, M., Cohen, L. B., & Chiarello, E. (2003). Six-month-old infants' categorization of containment spatial relations. *Child Development*, *74*, 1–15.
Cattel, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276.
Chiorri, C., Anselmi, P., & Robusto, E. (2009). Reverse items are not opposites of straightforward items. In U. Savardi (Ed.), *The perception and cognition of contraries* (pp. 295–328). Milano, IT: McGraw-Hill.
Clark, H. H. (1973). Space, time, semantics, and the child. In E. Timothy Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 27–62). London, UK: Academic Press.
Conrad, K. J., Wright, B. D., McKnight, P. E., McFall, M., Fontana, A., & Rosenheck, R. (2004). Comparing traditional and Rasch analyses of the Mississippi PTSD scale: Revealing limitations of reverse-scored items. *Journal of Applied Measurement*, *5*(1), 15–30.
Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge, UK: Cambridge University Press.
Cruse, D. A. (1986). *Lexical semantics*. Cambridge, UK: Cambridge University Press.

Cruse, D. A., & Togia, P. (1995). Towards a cognitive model of antonymy. *Journal of Lexicology*, *1*, 113–141.
Feather, N. T. (1980). Conservatism, acquiescence and the effects of sample heterogeneity. *Australian Journal of Psychology*, *32*, 11–16.
Fellbaum, C. (1995). Co-occurrence and antonymy. *International Journal of Lexicography*, *8*, 281–303.
Freeman, M. A., & Bordia, P. (2001). Assessing alternative models of individualism and collectivism: A confirmatory factor analysis. *European Journal of Personality*, *15*(2), 105–121.
Gärdenfors, P. (2000). *Conceptual spaces.* Cambridge — London, UK: The MIT Press.
Gärdenfors, P. (2007). Representing actions and functional properties in conceptual spaces. In T. Ziemke, J. Zlatev, & R. M. Frank (Eds.), *Body, language, and mind: Embodiment* (pp. 167–196). Berlin, Germany: Mouton de Gruyter.
Gepshtein, S. (2010). Two psychologies of perception and the prospect of their synthesis. *Philosophical Psychology*, *23*(2), 217–281.
González-Romá, V., Schaufeli, W. B., Bakker, A. B., & Lloret, S. (2006). Burnout and work engagement: Independent factors or opposite poles? *Journal of Vocational Behavior*, *68*, 65–174.
Gustafsson, R. L. (1980). Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology*, *33*, 205–233.
Harris, M. M., & Bladen, A. (1994). Wording effects in the measurement of role conflict and role ambiguity: A multitrait-mu/timethod analysis. *Journal of Management*, *20*(4), 887–901.
Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics*, *42*(1), 25–70.
Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, *21*, 967–988.
Hollins, M., Faldozski, R., Rao, S., & Young, F. (1993). Perceptual dimensions of tactile surface texture: A multidimensional scaling analysis. *Perception & Psychophysics*, *54*, 697–705.
Idaszak, J., & Drasgow, F. (1987). A revision of the job diagnostic survey: Elimination of a measurement artifact. *Journal of Applied Psychology*, *72*, 69–74.
Jones, S. (2002). *Antonymy: a corpus-based perspective.* London, UK: Routledge.
Kanizsa, G. (1979). *Organization in vision: Essays on Gestalt perception.* NY: Praeger.
Kelso, J. A. S., & Engstrom, D. A. (2006). *The complementary nature.* Cambridge MA: The MIT Press.
Kennedy, C., & McNally, L. (2005). Scale structure and the semantic typology of gradable predicates. *Language*, *81*(2), 345–381.
Kubovy, M. (2003). Phenomenology, psychological. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 579–586). Houndmills, Basingstoke, Hampshire, UK: Nature Publishing Group.
Kubovy, M., & Gepshtein, S. (2003). Grouping in space and in space-time: An exercise in phenomenological psychophysics. In R. Kimchi, M. Behrmann, & C. R. Olson (Eds.), *Perceptual organization in vision: Behavioral and neural perspectives* (pp. 45–86). Mahwah, NJ, USA: Erlbaum.
Kubzansky, L. D., Kubzansky, P. E., & Maselko, J. (2004). Optimism and pessimism in the context of health: Bipolar opposites or separate constructs? *Personality and Social Psychology Bulletin*, *30*(8), 943–956.
Lehrer, A., & Lehrer, K. (1982). Antonymy. *Linguistic and Philosophy*, *5*, 483–501.
Lorenzo-Seva, U., & Rodrigues-Fornells, A. (2006). Acquiescent responding in balanced multidimensional scales and exploratory factor analysis. *Psychometrika*, *71*(4), 769–777.
Marsh, H. W., & Myers, M. (1986). Masculinity, femininity, and androgyny: A methodological and theoretical critique. *Sex Roles*, *14*, 397–430.
McGee, G. W., Ferguson, C. E., & Seers, A. (1989). Role conflict and role ambiguity: Do the scales measure these two constructs? *Journal of Applied Psychology*, *74*, 815–818.
Miller, G. A. (1996). *The science of words.* New York, NY: McGraw-Hill.
Millis, K. K., & Neimeyer, R. A. (1990). A test of the dichotomy corollary: Propositions versus constructs as basic cognitive units. *International Journal of Personal Construct Psychology*, *3*, 167–181.
Molenaar, I. W. (1983). Some improved diagnostic for failure of the Rasch model. *Psychometrika*, *48*(1), 49–72.
Mosconi, G. (1967). Contributo all'analisi psicologica della qualità espressive [Contribution to the psychological analysis of perceptual qualities]. *Rivista di Psicologia*, *4*, 337–351.
Muehleisen, V. (1997). Antonymy and semantic range in English. *Doctoral dissertation in philosophy, field of linguistics, Evanston, Illinois.*
Murphy, L. (2003). *Semantic relations and the lexicon: Antonyms, synonyms and other semantic paradigms.* Cambridge, UK: Cambridge University Press.
Niemoller, K., & van Schuur, W. H. (1983). Stochastic models for unidimensional scaling: Mokken and Rasch. In D. McKay, N. Schofield, & P. Whitely (Eds.), *Data analysis and the social sciences* (pp. 120–170). London, UK: Pinter.
Paradis, C. (2001). Adjectives and boundedness. *Cognitive Linguistics*, *12*(1), 47–65.
Paradis, C. (2008). Configurations, construals and change: Expressions of DEGREE. *English Language and Linguistics*, *12*(2), 317–343.
Pecher, D., & Zwaan, R. A. (Eds.). (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking.* Cambridge, UK: Cambridge University Press.
Picard, D., Dacremont, C., Valentin, D., & Giboreau, A. (2003). Perceptual dimensions of tactile textures. *Acta Psychologica*, *114*, 165–184.
Pinquart, M., Fröhlich, C., & Silbereisen, R. K. (2007). Optimism, pessimism, and change of psychological well-being in cancer patients. *Psychology, Health & Medicine*, *12*(4), 421–432.
Pinna, B. (2010). What comes before psychophysics? The problem of "what" we perceive and the phenomenological exploration of new effects. *Seeing and Perceiving*, *23*, 463–481.
Podsakoff, P. M., MacKenzie, S. B., Lee, J. -Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903.
Quinn, P. C. (2005). Developmental constraints on the representation of spatial relation information: Evidence from preverbal infants. In L. A. Carlson, & E. van der Zee (Eds.), *Functional features in language and space: Insights from perception, categorization, and development* (pp. 293–309). New York, NY: Oxford University Press.
Rasch, G. (1960; ed. 1980). *Probabilistic models for some intelligence and attainment tests.* Chicago, USA: The University of Chicago Press.
Ray, J. J. (1983). Reviving the problem of acquiescent response bias. *Journal of Social Psychology*, *121*, 81–96.
Ray, J. J. (1985). Acquiescence and response skewness in scale construction: A paradox. *Personality and Individual Differences*, *6*(5), 655–656.
Rodebaugh, T. L., Woods, C. M., & Heimberg, R. G. (2007). The reverse of social anxiety is not always the opposite: The reverse-scored items of the social interaction anxiety scale do not belong. *Behavior Therapy*, *38*, 192–206.
Rost, J. (1982). An unconditional likelihood ratio for testing homogeneity in the Rasch model. *Education Research and Perspectives*, *9*, 7–17.
Russell, J. A., & Carroll, J. M. (1999). The phoenix of bipolarity: Reply to Watson and Tellegen. *Psychological Bulletin*, *125*(5), 611–617.
Savardi, U., Bianchi, I., & Burro, R. (2009). From opposites to dimensions: Filling in the gaps. In U. Savardi (Ed.), *The perception and cognition of contraries* (pp. 275–294). Milano, IT: McGraw-Hill.
Schermelleh-Engel, K., Moosbrugger, H., & Muller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and goodness-of-fit models. *Methods of Psychological Research — Online*, *8*(2), 23–74.
Schimmack, U. (2001). Pleasure, displeasure, and mixed feelings? Are semantic opposites mutually exclusive? *Cognition and Emotion*, *15*, 81–97.
Schmicking, D., & Gallagher, S. (2010). *Handbook of phenomenology and cognitive science.* Dordrecht: Springer Verlag.
Tracy, L., & Johnson, T. W. (1981). What do the role conflict and role ambiguity scales measure? *Journal of Applied Psychology*, *66*, 464–469.
Verheyen, S., Hampton, J. A., & Storms, G. (2010). A probabilistic threshold model: Analyzing semantic categorization data with the Rasch model. *Acta Psychologica*, *135*, 216–225.
Vidotto, G., Robusto, E., & Zambianchi, E. (1996). I modelli simple logistic e rating scale nella determinazione del punto di eguaglianza soggettivo: una nuova prospettiva per il metodo degli stimoli costanti [Simple logistic and rating scale models in defining the point of subjective equality: A new perspective for the method of constant stimuli]. *TPM (Testing Psicometria Metodologia)*, *3*(4), 227–234.
Watson, D., & Tellegen, A. (1999). Issues in the dimensional structure of affect. Effects of descriptors, measurement error, and response formats: Comment on Russell and Carroll (1999). *Psychological Bulletin*, *125*, 601–610.
van den Wollenberg, A. L. (1982). Two new test statistics for Rasch model. *Psychometrika*, *47*(2), 123–140.
Wood, R. (1978). Fitting the Rasch model: A heady tale. *British Journal of Mathematical and Statistical Psychology*, *31*, 27–32.
Wright, B. D. (1980). Afterwords. In G. Rasch (Ed.), *Probabilistic models for some intelligence and attainment tests.* Chicago, IL: University of Chicago Press.
Wright, B. D. (1984). Despair and hope for educational measurement. *Contemporary Education Review*, *3*(1), 281–288.
Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis.* Chicago, IL: University Press Chicago.
Wright, B. D., & Masters, G. N. (1981). *The measurement of knowledge and attitude (research memorandum no. 30).* Chicago, IL: University of Chicago, MESA Psychometric Laboratory.
Wright, B. D., & Stone, M. H. (1979). *Best test design.* Chicago, IL: MESA Press.
Yorke, M. (2001). Bipolarity… or not? Some conceptual problems relating to bipolar rating scales. *British Educational Research Journal*, *27*(2), 171–186.
Zwinderman, A. H. (1995). Pairwise estimation in the Rasch models. *Applied Psychological Measurement*, *19*(4), 369–375.