Synthese (2017) 194:4011–4037
DOI 10.1007/s11229-016-1123-z

A case against convexity in conceptual spaces

José V. Hernández-Conde[1]

Received: 26 October 2015 / Accepted: 11 May 2016 / Published online: 28 May 2016

Abstract The notion of conceptual space, proposed by Gärdenfors as a framework for the representation of concepts and knowledge, has been highly influential over the last decade or so. One of the main theses involved in this approach is that the conceptual regions associated with properties, concepts, verbs, etc. are convex. The aim of this paper is to show that such a constraint—that of the convexity of the geometry of conceptual regions—is problematic; both from a theoretical perspective and with regard to the inner workings of the theory itself. On the one hand, all the arguments provided in favor of the convexity of conceptual regions rest on controversial assumptions. Additionally, his argument in support of a Euclidean metric, based on the integral character of conceptual dimensions, is weak, and under non-Euclidean metrics the structure of regions may be non-convex. Furthermore, even if the metric were Euclidean, the convexity constraint could be not satisfied if concepts were differentially weighted. On the other hand, Gärdenfors' convexity constraint is brought into question by the own inner workings of conceptual spaces because: (i) some of the allegedly convex properties of concepts are not convex; (ii) the conceptual regions resulting from the combination of convex properties can be non-convex; (iii) convex regions may co-vary in non-convex ways; and (iv) his definition of verbs is incompatible with a definition of properties in terms of convex regions. Therefore, the mandatory character of the convexity requirement for regions in a conceptual space theory should be reconsidered.

Keywords Conceptual spaces · Convexity constraint · Prototype theory · Cognitive semantics

✉ José V. Hernández-Conde
jhercon@gmail.com

1 Department of Linguistics and Basque Studies, Faculty of Arts, University of the Basque Country, Paseo de la Universidad 5, 01006 Vitoria-Gasteiz, Álava, Spain

1 Introduction

The notion of conceptual space, proposed by Gärdenfors (2000) as a framework for the representation of concepts and knowledge, has been highly influential over the last fifteen years. Since his initial proposal, Gärdenfors (2014) has tried to extend the approach both to the modeling of actions and events, and to the semantics of verbs, prepositions and adverbs. One of the basic theses of his approach is that the conceptual regions associated with properties, concepts (or object categories), verb meanings, etc. are convex. The aim of this work is to show that such a constraint, that of the geometrical convexity of conceptual regions, is problematic; not only from a theoretical perspective, but also with regard to the inner workings of the theory itself.

In Sect. 2, after this brief introduction, I expound the main features of Gärdenfors' theory of conceptual spaces, focusing on his definitions of property and concept. At the same time I aim to explain clearly the role played by the convexity constraint in the theory, as opposed to other possible criteria that could be imposed on the geometry of regions. In Sect. 3 I recap how the notion of similarity is characterized within a geometrical approach; and I introduce the distinction between standard and non-standard distances. There, and after introducing the notion of Voronoi partition, I hold that Gärdenfors is tacitly committed to the thesis that conceptual regions are the cells resulting from a Voronoi tessellation starting with a set of prototypes.

Nevertheless, the regions resulting from a Voronoi tessellation are only convex under very specific assumptions, namely, if the metric is Euclidean, and if distances are not differently weighted for each particular concept. Therefore, Gärdenfors may try to defend the convexity constraint through two different strategies: either arguing directly in favor of the convexity of the conceptual regions; and/or arguing for the assumptions which guarantee that the regions produced by a Voronoi tessellation are convex. I will try to show that: [I] Gärdenfors' direct arguments in favor of convexity are not conclusive. [II] The assumption of a Euclidean metric is not guaranteed. [III] It is not implausible that the distances to the prototypes of distinct concepts are differently weighted.

Section 4 focuses on criticism of the major cognitive reasons for the convexity of conceptual regions: (a) co-implication with the prototype theory of concepts; (b) cognitive economy; (c) its perceptual foundations; and (d) effectiveness of communication. I aim to show that none of them compels us to accept the convexity constraint as compulsory.

Notwithstanding that, if the metric underlying conceptual spaces were the standard Euclidean metric (that is, if distances were Euclidean and non-weighted), then the convexity of regions would be guaranteed. With regard to this point, Gärdenfors argues that, for the case of integral dimensions, the Euclidean metric fits the empirical data better than the city-block metric. So, given that his definitions of property and concept are for domains constituted by sets of integral dimensions, it is possible to conclude that the metric of conceptual spaces is the Euclidean one. Section 5 brings into question the empirical evidence that supports such an argument; evidence supposedly in favor of the relation of mutual dependency between integral dimensions and the Euclidean metric. By questioning the evidence in this way, I intend to show that the metric underlying conceptual spaces can be non-Euclidean, and in that case the convexity constraint on regions does not hold, as I indicate in Sect. 6. After that, in Sect. 7, I allege that, even if conceptual spaces do function with a Euclidean framework, conceptual regions might be non-convex if distances of comparison in categorizations are weighted differently.

Finally, Sect. 8 is devoted to proving that the convexity requirement is brought into question by the very characterization of the inner workings of Gärdenfors' conceptual space theory itself. From all these arguments, I conclude (Sect. 9) that the mandatory character of the convexity constraint should be rethought, perhaps in favor of a weaker (and non-mandatory) criterion for the geometry of regions.

2 Conceptual spaces and the convexity constraint

Gärdenfors proposal consists of a non-connectionist theory of conceptual spaces based on the notion of similarity. In general terms, a similarity space theory of concepts can be described by the following fundamental thesis (Gauker 2007): the mind is a representational hyperspace within which (a) *dimensions* represent ways in which objects can differ, (b) *points* represent objects, (c) *regions* represent concepts, and (d) *distances* are inversely proportional to *similarities* (between objects or concepts). Consequently, an object will belong to a concept if and only if its values in every dimension of that similarity space produce an n-tuple that lies inside the region associated with that concept.

2.1 Gärdenfors' conceptual spaces

Nonetheless, there are important differences between the conceptual space theory proposed by Gärdenfors (2000) and this general framework. Firstly, (natural) *properties* are convex regions of a given domain (Criterion P), and are typically associated with the meaning of adjectives. Secondly, (natural[1]) *concepts* are bundles of properties (or, alternatively, sets of convex regions) in a number of domains, together with the salience weights of those domains and information concerning how their regions are correlated (Criterion C); and they typically represent the meaning of nouns. Finally, in this framework, the notion of *domain* is critical: it is defined as a set of integral[2] dimensions that are separable from all other dimensions.

In accordance with this general scheme, concepts are a result of the division of the similarity space into convex regions (constituted by the sets of points representing those objects that exhibit the sensory properties characteristic of such regions). The convex regions are identified by Gärdenfors precisely with concepts.

1 Although these two definitions are for *natural* properties and concepts, as a matter of fact Gärdenfors applies them almost universally: he does not distinguish between *natural* and *non-natural* properties or concepts (except in order to discriminate artificial non-convex properties or concepts, such as those associated with Goodman's term *grue*: *green* before a given date and *blue* after that date).

2 Integral dimensions are those processed in a holistic and unanalyzable way, where the assignation of a value to a particular dimension requires a value to be given to the others. If dimensions are not integral, then they are separable. Gärdenfors' main domain example is the case of *color*, which would be constituted of three dimensions: *hue*, *intensity* and *brightness*.

Lastly, in his most recent work, Gärdenfors and his collaborators have tried to extend this basic framework from the cases of properties and concepts (or, alternatively, from adjectives and nouns) to the representation of states, changes, actions and events; through these, to the semantics of verbs, adverbs and prepositions; and ultimately to apply it to the case of human communication (Gärdenfors and Warglien 2012, 2013; Warglien et al. 2012; Gärdenfors 2014). Very briefly sketched, Gärdenfors takes as a starting point the thesis that *verbs* typically represent dynamic properties of objects; are parts of *events*; and involve *actions* which are constituted by *forces* (commonly exerted by agents). In virtue of this, his proposal for verbs consists, in effect, of a holistic model of actions, forces, events and verbs, characterized by means of conceptual spaces. Within such a framework, *verbs* denote changes in properties (that is, movements in the representation of objects or concepts within the conceptual space), and they both refer to and are represented by convex regions of vectors.

2.2 The convexity constraint

As is evident from the previous subsection, the requirement for the convexity of regions runs through all the conceptual space theory defended by Gärdenfors. Not only properties and concepts (or object categories), but also the semantics of verbs, adverbs and prepositions (Gärdenfors 2014) are conceived and represented within his theory by convex regions.

The convexity requirement can be thought of as a generalized definition of the conception of a natural kind as a qualitatively spherical region, expounded by Quine (1969) in his discussion of the definition of natural kinds in terms of (comparative) similarity. Gärdenfors' aim is to characterize the geometrical form of natural properties and concepts, when acting as (optimal) evolutionary tools in tasks such as problem-solving, memorizing, planning, communicating, etc. To that end, he distinguishes three possible criteria which could constrain the geometry of a region (see Fig. 1):

- *connectedness* constraint: it must be possible to reach every point in the region from every other point by following a continuous path consisting only of points belonging to the region;
- *star-shapedness* constraint (with respect to a point P): for every point x in the region, all the points between[3] x and P must belong to the same region;
- *convexity* constraint: the region must satisfy the star-shapedness constraint with respect to all the points in the region, that is, for every two points in the region, all the points between them must also belong to the same region.

The strength of these three criteria increases in order: every star-shaped region is a connected region and (trivially) every convex region is a star-shaped region.

[3] Although the second and third constraints require the definition of *betweenness* [an axiomatic definition of which can be found in Gärdenfors (2000, p. 15)], I will not discuss that topic in this paper.

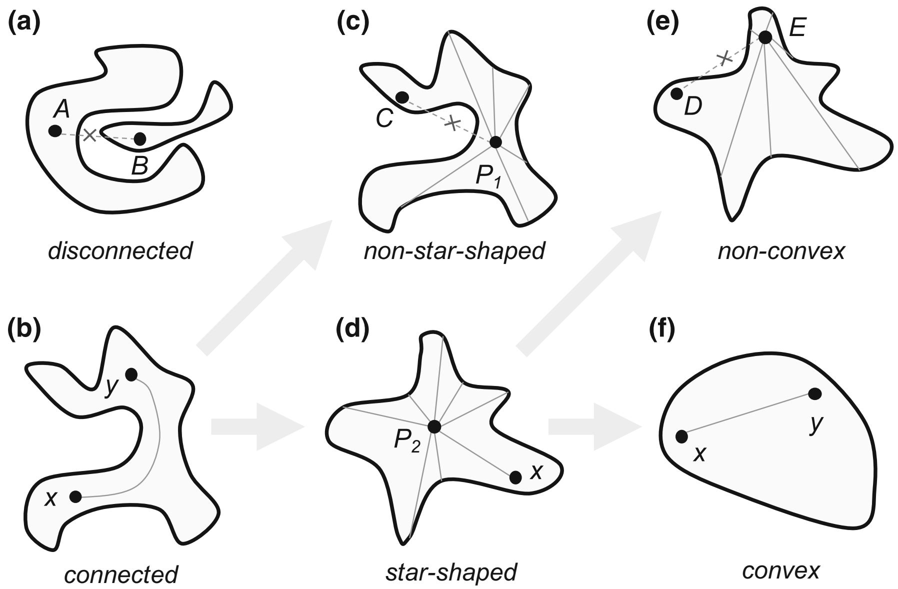

Fig. 1 Representation of the three different criteria for the geometry of conceptual regions: *connectedness*, *star-shapedness* and *convexity*. Paths containing exclusively points belonging to the considered regions are represented by *solid lines*. Paths also containing points outside the regions are represented by *dashed lines*. The problematic points not belonging to the regions are represented by *crosses*. **a** Representation of a *disconnected* region R_1 where the point A is not reachable from the point B following a continuous path of points belonging to R_1. **b** Representation of a *connected* region R_2 where every point x in R_2 is reachable from every other point y in R_2, following a continuous path of points belonging to R_2. **c** Representation of a *connected* but *non-star-shaped* region R_2 (the same as in graph **b**), where there is no point with respect to which R_2 satisfies the star-shapedness constraint (for instance, P_1 cannot be that point because between P_1 and C there are points not belonging to R_2). **d** Representation of a *connected* and *star-shaped* region R_3, where there is a point P_2 in R_3 such that, for every point x in R_3, all the points between P_2 and x also belong to R_3. **e** Representation of a *star-shaped* but *non-convex* region R_3 (the same as in graph **d**), where there are points D and E in R_3 such that not all the points between them also belong to R_3. **f** Representation of a *star-shaped* and *convex* region R_4, where for every two points x and y in R_4, all the points between x and y also belong to R_4

3 Similarity measures and Voronoi diagrams

3.1 Geometric similarity measures

There are four main approaches to characterizing the notion of similarity (Goldstone and Son 2005): geometrical, feature based, alignment based and transformational. Here, in the spirit of Gärdenfors' conceptual spaces, I focus on geometric characterizations of similarity. Models based on this approach define similarity as a measure that is inversely proportional to distance, which is usually determined according to a Minkowski metric. Let us remember the expression for the distance (in a generic Minkowski metric) between two objects (and/or prototypes of concepts) A and B located within an n-dimensional space, where $X_i^{[Y]}$ represents the value of the i-th dimension associated with the concept Y:

$$d(A, B) = \left(\sum_{i=1}^{n} \left| X_i^{[A]} - X_i^{[B]} \right|^p \right)^{1/p}$$

The value of the parameter p determines the type of metric and distance: if $p = 1$, they are called Manhattan (or city-block); when $p = 2$, they are called Euclidean.

This expression corresponds to *ordinary* Minkowski distances. Nevertheless, these distances can be weighted differently according to various criteria. For example, in the case of conceptual spaces, the weights could be a function of the number of particular cases (instances or examples) on which a given concept is based. In such a case, the distance-of-comparison in categorizations of a certain object, O, with respect to a particular concept, C_i (represented within the conceptual space by the prototype P_{C_i}), referred to as $d_{Ci}(O, P_{C_i})$, could be expressed, under a multiplicatively weighted scheme,[4] as follows[5]:

$$d_{Ci}(O, P_{C_i}) = w_i d(O, P_{C_i})$$

where w_i represents the weight assigned to that concept. In Sect. 7 below, I will show the implications of non-standard weighting with regard to the convexity requirement.

3.2 Voronoi tessellations of a conceptual space

Once assumed a geometric similarity measure, Gärdenfors' proposal can be characterized by means of Voronoi diagrams, inasmuch as concepts and properties can be conceived as the cells resulting from a Voronoi tessellation of the conceptual space.

A Voronoi diagram is a partition of an n-dimensional space into regions, based on the distances between each point and the points belonging to a particular subset G of that n-dimensional space. The points belonging to G are commonly called *seeds* or *generators*, and in Gärdenfors' theory those points are the *prototypes* of concepts. The general idea is that for each generator g_i there exists a region constituted by those points nearest to g_i than to any other seed belonging to G. The points equidistant from their two closest generators will constitute the boundaries of regions. Thus, for example, in the case of a standard Euclidean metric where both concepts and dimensions were equally weighted, the boundaries of regions would be determined by the bisectors of the segments connecting each pair of generators.

On my view, Gärdenfors is strongly committed to the thesis that the shapes and boundaries of conceptual regions are produced by a Voronoi tessellation of the conceptual hyperspace, whose input are the prototypes of the relevant concepts. Regarding this, it could be objected that Gärdenfors never expresses openly his commitment to the thesis that conceptual regions are the cells resulting from a Voronoi tessellation starting with a set of prototypes (THESIS V). Nevertheless, although that is true, it is

[4] For a detailed review of approaches to weighting that are distinct from the multiplicative one, see Okabe et al. (1992, pp. 119–134).

[5] *Ordinary distances* used simply to be called *distances* (or *standard distances*), and that is what I will do in this work; conversely, I will use the terms *weighted distance* and *non-standard distance* indistinctly.

also clear that he accepts THESIS V tacitly when he explains how significant elements of his theory work. For instance, in *The geometry of meaning* (2014), his explanations with regard to conceptual learning (Gärdenfors 2014, p. 42), conceptual change (ib., p. 43), categorization (ib., pp. 27–28), communication and language (ib., pp. 274–275), and vagueness (ib., pp. 45–46) are all of them based on the assumption of THESIS V.

In fact, if Gärdenfors considered that conceptual regions might be something different to cells resulting from a Voronoi tessellation, then he should have provided a distinct kind of explanation. More specifically, if he considered that Voronoi partitions are only one possible way by which his conceptual space theory could be articulated, he ought to have explained those phenomena for a characterization K of his theory such that Voronoi tessellations were a particular case of K. The fact that he provides no explanations other than the ones based on THESIS V is indicative of his strong (although tacit) commitment to that thesis. In consequence, throughout all my paper I will assume that Gärdenfors accepts THESIS V.

4 Gärdenfors' arguments for the convexity constraint

In his work, Gärdenfors does not provide any 'definitive' argument in favor of the convexity constraint, but he offers a series of reasons that suggests a high degree of plausibility for it. My aim in this section is to show that none of those arguments is compelling: we do not have to accept convexity as a mandatory requirement for the geometry of regions.

4.1 Mutual dependence with the prototype theory

One of the six basic principles that Gärdenfors considers to be embodied by the cognitive approach to semantics is that concepts show prototypical effects which cannot be explained from the standpoint of classical theories of concepts. In fact, one of the main advantages of Gärdenfors' approach is that his conceptual spaces provide a natural explanation of prototypical effects for many concepts[6] (Rosch 1978; Lakoff 1987). Let us see why.

According to the prototype theory, concepts are prototypes, that is, representations whose structure encodes information about the properties that their members tend to have. However, there are different ways in which the prototype theory can be articulated (Smith and Medin 1981): (a) *featural models*: an object is classified under a given category if it possesses a sufficient number of the properties associated to that concept; (b) *dimensional models*: an object is classified under a given category if it possesses to some degree a sufficient number of those properties. (Gärdenfors' conceptual spaces are a particular type of dimensional models, where distances in a certain dimension are indicative of the degree-of-possession of the properties associated to that dimension.) In both cases an object will be categorized or not under a particular concept in function

[6] Prototypical effects are associated with the fact that some members of a category are considered more representative of it than others. For instance, *robins* are considered more representative of the BIRD category than *eagles,* and *eagles* more than *chickens*.

of the similarity between the object and the prototype of that concept. Their similarity will be determined by virtue of their shared properties (either possessed, or possessed to some degree, depending on whether the model is featural or dimensional). In consequence, the more prototypical a member of a category is: (i) the more attributes it shares with the other members of the category; and (ii) the fewer features it shares with the members of other categories (Rosch and Mervis 1975).

And, with regard to the question of how the prototypes of concepts are acquired, it is reasonable to think that those prototypes are the result of a process of maximization of similarities (or, alternatively, minimization of distances) between the evaluated objects, and the tentative prototype of a particular category. The final set of prototypes will be the one which maximizes intra-group similarity and minimizes inter-group similarity. Therefore, the prototype of a concept arises from the generalization of the properties of the objects chosen as tentative members of a given category. Hence, prototypes are those members (whether real, or not[7]) of a category that best reflect the similarity structure of the category as a whole. Next, a Voronoi partition of the conceptual hyperspace can be produced, whose input would be the prototypes of the relevant concepts.

In addition, in order to categorize a new object under a particular concept, that object will only need to be evaluated with respect to the prototypes of the relevant categories, which makes the categorization process very efficient (with no need to resort to the shapes or boundaries of regions).

Based on this, it is possible to say that Gärdenfors' conceptual space theory provides a natural explanation of the typicality effects we expect a prototype theory of concepts to exhibit, given that: (i) in a conceptual space the centers of gravity (or mass centroids) of the regions associated with each concept can be identified with the most representative members endorsed by the prototype theory of concepts; and (ii) prototypicality can be characterized as a measure that is inversely proportional to the distance (of an object) from those centers of gravity.[8]

---

[7] That is, with or without real instances of them.

[8] Here it must be said that, within the conceptual space framework, prototypicality effects in a region can happen regarding two main different referents:

[I] With regard to the *centers of gravity*, or mass centroids, of the regions: this option is quite straightforward (and the one adopted by Gärdenfors), because the center of gravity of a convex region is the most typical member of that region. This alternative also implies a strong commitment to the literal reading that properties and concepts are represented by convex regions (and not merely by their prototypes). This is so because, in order to determine the center of gravity it is necessary to evaluate the whole region associated to the concept; and the region must be convex so that its center of gravity belongs to the region.

[II] With regard to the *prototypes*, or Voronoi generators, of the regions: this alternative has a crucial difference with respect to option [I]: the prototype is the most typical member, not of the region resulting from a Voronoi tessellation based on that prototype, but of the instances (particular cases or examples) from which the prototype of that concept was determined (for example, by means of a cluster analysis). This approach is more compatible with the principle of cognitive economy than the first, because in alterative [I] the evaluation of typicality requires (i) calculating or storing the whole conceptual region, and (ii) determining therefrom the gravity center of that region. In contrast, in option [II] typicality evaluation only needs the determination of the distance to the prototype.

With regard to this first claim, Gärdenfors defends the notion that those who adopt the prototype theory of concepts should expect a representation of concepts and properties as convex regions; and contrariwise, that if concepts are characterized as convex regions, then prototypical effects should be expected (Gärdenfors 2000, pp. 86–87; 2014, pp. 26–27).

It is my view that this is the main argument offered by Gärdenfors in support of the convex geometry of regions. However, neither of the two assertions constitutes a reason in favor of the convexity requirement, given that both of them could also be applied to a star-shaped region resulting from a Voronoi tessellation, as I now explain.

[A] *If properties and concepts were defined as star-shaped regions* (produced by a Voronoi tessellation) *then prototypical effects would also be expected*: in this case the typicality of an object with regard to a given category is also a function of the distance between the point representing that object and the prototype of its category.

[B] *The only thing that should be expected by a consistent prototype theorist is the star-shapedness of conceptual regions*: a prototype theorist should expect that if an object belongs to a certain category, then all the objects with the same proportional distances from the prototype but more similar to it (that is, all the objects between the object under consideration and the prototype), should also belong to that category. This is exactly what happens under the star-shapedness constraint.[9]

Thus, Gärdenfors' alleged mutual dependence between the prototype theory and the convexity of regions (within a conceptual space approach articulated by means of Voronoi tessellations), also takes place between the theory of prototypes and the star-shapedness of regions. Consequently, such a relationship cannot be a crucial reason in favor of the convexity constraint.[10]

Footnote 8 Continued
And, even though Gärdenfors chooses option [I], both approaches are equally acceptable in order to explain typicality effects in conceptual regions. In fact, option [II] is the one chosen by me when I argue below that the star-shaped regions which result from a Voronoi tessellation can also explain prototypicality.

[9] The argument which connects the prototype theory (articulated by means of Voronoi tessellations) with the star-shapedness of regions can be summed up as follows:

PREMISE 1: If an object, O, belongs to a concept, C (characterized by a prototype P), this entails that the ball $B(O, OP)$, centered at O and with radius OP, does not contain any other prototype distinct from P. [PREMISE 1 is equivalent to the thesis that concepts are the result of a Voronoi tessellation (THESIS V)].

PREMISE 2: Minkowski metric with $p \geq 1$ and non-weighted prototypes.

CONCLUSION: Conceptual regions are star-shaped.

The general idea is that for every object A, between O and P, it is possible to prove that the ball $B(A, AP)$ is included within $B(O, OP)$. Therefore, P is the nearest prototype to A; that is, the object A also belongs to C and, in consequence, conceptual regions are star-shaped. [For the specific details of this proof see Lemma 5 in Lee (1980, p. 608)].

[10] Withal, this should not be seen as a defense of a mandatory star-shapedness constraint, because it could be accepted the different weighting of prototypes, and in that case the conceptual regions might not be star-shaped (as shown in Fig. 3b).

4.2 Cognitive economy

When Gärdenfors originally defined properties in terms of convex regions, he mainly based his decision on the argument provided by Shepard (1987, p. 1319). Shepard argued that evolution would have led to consequential regions (in our psychological space) in a way such that the boundaries of those regions were not oddly shaped. Next Gärdenfors (2000, p. 70) maintained that such an evolutionary preference could be supported by a principle of *cognitive economy* in terms of memory, learning and processing.

However, the cognitive economy argument depends on the assumption that the handling of convex sets of points requires less memory, learning and processing resources than the handling of regions with capricious forms. In this case, Gärdenfors' argument can be structured as follows.

(i) Properties and concepts are determined by convex regions within a conceptual hyperspace. (CRITERIA P and C)
(ii) Those convex regions can be the result of a Voronoi tessellation starting with a set of prototypes.[11] (THESIS V)
(iii) A Voronoi tessellation could support, in a cognitively efficient way, psychological processes such as concept learning, categorization, communication and language.[12] Additionally, Voronoi tessellations can also explain other phenomena, such as conceptual change and vagueness.
(iv) Therefore, the handling of convex regions could explain cognitive efficiency in all those tasks and processes.

The problem is that, as shown in Sects. 6 and 7 below, the conceptual regions can be non-convex and yet compatible with a Voronoi tessellation. Moreover, given that the reasons offered by Gärdenfors are not exclusive to convex regions, an independent argument for the greater cognitive efficiency of handling convex regions (over non-convex ones) is still required.

I entirely support Gärdenfors' aim of basing the conceptual space theory on reasons related to cognitive economy; in fact, I think that most of his arguments are thoroughly valid with regard to such a general point. That is, a conceptual space theory is highly efficient from a cognitive point of view, given that:

- only prototype locations have to be memorized;
- in categorization tasks the only distances evaluated would be those between the evaluated objects and the prototypes associated with the categories considered; and
- concepts could be learned from a very small number of particular examples.

[11] In order that the resulting regions may be convex, the Voronoi tessellation will have to meet some conditions (Euclidean metric, and non-weighted prototypes).

[12] In Gärdenfors' words: "The [Voronoi] tessellation mechanism provides important clues to the cognitive economy of concept learning. If the categorization of each point in a space had to be memorized, this would put absurd demands on human memory. However, if the partitioning of a space into categories is based on a Voronoi tessellation, only the relative positions of the prototypes need to be remembered" (Gärdenfors 2014, pp. 27–28). Gärdenfors makes use of similar arguments for the cognitive efficiency of applying Voronoi tessellations in his explanation of how human communication works (ib., pp. 274–275).

Nonetheless, all these facts are common to every conceptual space theory which assumes that concepts are represented by regions resulting from a Voronoi tessellation, independently of the geometrical structure (convex or non-convex) of those conceptual regions. Consequently, cognitive efficiency cannot be a crucial reason to support the convexity criterion.

4.3 Perceptual foundation

Gärdenfors usually contends that many perceptually grounded domains, such as color, taste, vowels, etc., are convex, based on evidence in favor of the convexity of the regions associated with numerous typical properties of all those domains (Fairbanks and Grubb 1961; Sivik and Taft 1994). The *color* domain, however, seems to be his preferred example of integral dimensions. In the case of color, the problem is that the work that Gärdenfors refers to as evidence is entirely associated with sensory dimensions; there is no guarantee that things work in the very same way in non-perceptual domains. This last point is explicitly recognized by Gärdenfors (2014, p. 137) when he acknowledges that the evidence (mainly associated with the *color* domain) does not provide automatic support for the convexity constraint in other domains.

4.4 Effectiveness of communication

One of the most recent arguments offered by Gärdenfors (2014, p. 26) for the convexity requirement is that the convexity of conceptual regions is decisive in effective communication. In this case Gärdenfors argues that Jäger's research has shown that in language, convex regions are a result of cultural evolution (Jäger 2007). However, Jäger's work does not constitute an argument either in favor of or against the geometrical structure of conceptual spaces, given that he assumes the standard[13] Euclidean metric (Jäger 2007, p. 554); so semantic categories have to be convex (see below for discussion of this point). Due to this, when Gärdenfors cites Jäger's research in support of his thesis that conceptual regions are convex, he falls into *petition principii*: since Jäger starts from the standard Euclidean metric, his results simply do not disprove the convexity constraint thesis; but they cannot confirm it.

To sum up, on the one hand, under the assumption of Thesis V neither the mutual dependence with the prototype theory of concepts, nor the economy cognitive argument are definitive reasons in favor of convexity. That was so because those arguments are equally valid for all the conceptual regions resulting from a Voronoi tessellation, independently that their shapes are convex or not. On the other hand, none of the two other arguments for convexity (perceptual foundation and effectiveness of communication) are convincing enough to accept that concepts have to be convex.

[13] According to the notions of standard and non-standard distances given in footnote 5 above.

5 Integral dimensions, Euclidean metric, and convexity

As stated above, Gärdenfors' conceptual spaces rest on the assumption that regions are convex and, if concepts are not differently weighted, this convexity is guaranteed under a Euclidean metric (Gärdenfors 2000, p. 88; Okabe et al. 1992, p. 57). By virtue of this, the theory requires that the metric underlying our psychological space is Euclidean. On this occasion, the main argument in favor of a Euclidean metric is that in the case of integral dimensions, a Euclidean metric fits the empirical data better than a city-block metric (the latter would be more appropriate in the case of separable dimensions). Furthermore, given that Gärdenfors' definitions of property and concept are for domains constituted by sets of integral dimensions, it is possible to conclude that the conceptual spaces underlying them function with a Euclidean metric and, consequently, that their associated regions are convex.

However, this argument presents several problems, mainly due to the alleged mutual dependency relationship between integral dimensions and the (standard) Euclidean metric.

- First, Gärdenfors adduces a sort of co-implication between integral domains and the Euclidean metric: "If the Euclidean metric fits the data best, the dimensions are classified as integral; (...) when the dimensions are integral, the dissimilarity is determined by both dimensions taken together, which motivates a Euclidean metric" (Gärdenfors 2000, p. 25). The first implication is true, because if the metric is not city-block, the dimensions are non-separable (that is, they are integral). Nonetheless, the second conditional is false, inasmuch as the non-separability of dimensions does not necessarily imply that the metric is Euclidean.[14]
  Thus, the Euclidean character of the metric structure cannot be based on the integral character of domains. Gärdenfors assumes that dimensions are integral, but that is not sufficient to guarantee that the metric is Euclidean. In consequence, the Euclidean structure of a metric space needs empirical evidence independent from the one associated to the non-separability of its constitutive dimensions.[15]
- Second, and even assuming (as Gärdenfors does) that the co-implication between integral dimensions and the Euclidean metric were the case, the empirical evidence referred to in favor of the integral or separable character of a particular set

[14] On the one hand, the city-block metric is accepted when dimensions are separable because, in that case, the dimensions are the most meaningful element: they contribute independently to the total distance, and must remain invariant (unrotated) to keep the same conceptual space structure (Garner 1974, p. 119). On the other hand, the Euclidean metric is proposed when dimensions are non-analyzable because, in this case, distance (which remains the same for all rotations of axes) is the most relevant element.
However, it is possible that [i] although the dimensions did not contribute independently to the value of distances (and, therefore, distance were the most meaningful element); [ii] it happened that, despite of [i], the conceptual space structure were not irrelevant, and distances were not invariant under rotations of dimensions. In that case, dimensions would be non-separable (due to [i]) but, at the same time, they would not be secondary (by virtue of [ii]). Therefore, the non-separability of dimensions does not imply a Euclidean structure of the underlying conceptual space, whose metric could be non-Euclidean (for instance, with parameter p equal to 1.7 or 3) and still able to explain the non-separability of domains.

[15] Thanks to an anonymous reviewer whose comments led me to rethink this point.

of dimensions is tied to perceptual domains,[16] such as *color*, *sound*, *size*, *shape*, etc. (Garner 1974; Maddox 1992; Melara 1992). All that work faces a threefold difficulty, when taken as evidence in favor of Gärdenfors' theses, as I now explain.

[A] All the work is based on classification experiments and judgments of similarity at a conscious level, in which the integral character of dimensions (and, in consequence, the Euclidean character of the metric) could depend on how those classifications and judgments are consciously carried out, and not on the geometrical structure of the perceptual space.

[B] The experiments were developed over a small number of perceptual domains, so accepting them as evidence of the geometry of conceptual spaces requires the assumption that the behavior of the metric structure is the same across all perceptual and conceptual domains.[17] That is, it is necessary to assume that such behavior extends not only from the perceptual domains studied to all other perceptual domains, but also to all conceptual domains (in general not related to any of the perceptual domains studied), which might not be the case.

[C] This kind of work is used to contrast Euclidean and city-block metrics, and shows that the former fits integral sets of dimensions better, while the latter provides a better fit when the dimensions are separable. In the case in hand, however, the problem is that both metrics provide *good* fits, but not *perfect* fits. This ultimately means that the best metric is neither the Euclidean nor the city-block one; but is something between the two.

For example, in Handel and Imai (1972, p. 110) the optimal parameter p for integral dimensions in a general Minkowski metric is 1.7, which may be acceptable as reasonably close to a Euclidean space, but with non-convex regions[18] (given that their convexity would require a value of p equal or much closer to 2).

Therefore, what can be derived from this work is not that the (standard) Euclidean metric is warranted for integral dimensions, but only that the expected metric for integral domains will be closer to the (standard) Euclidean metric than to the (standard) city-block metric.

To sum up, all this evidence appears to be controversial; both that supporting the integral character of conceptual dimensions, and that which allegedly backs up the relationship between the integral character of dimensions and the Euclidean metric. The consequence is that, in both cases, the underlying metric could be non-Euclidean and, hence, conceptual regions could be non-convex.

[16] As happened with the perceptual foundation argument in favor of the convexity constraint (see Sect. 4).

[17] This behavior of the metric structure can be summed up as follows: separable dimensions are better characterized by a city-block metric, while the Euclidean metric is the best for integral dimensions.

[18] See Sect. 6 below for the meaning of the p parameter within the standard Minkowski metric. That section also contains a chart (Fig. 2) which shows that for a parameter p equal to 1.7 the conceptual regions are not convex.

6 Conceptual spaces under a non-Euclidean metric

Merely from attending to the basic requirements of a similarity space theory of concepts, it can be seen that the convexity constraint is unnecessary: nothing in the general conception of this kind of theories demands a Euclidean metric, and under a non-Euclidean metric the conceptual regions resulting from a Voronoi tessellation can be non-convex. Nonetheless, a constant throughout all of Gärdenfors' work is that he explicitly adopts a Euclidean metric which apparently guarantees the convexity of the conceptual regions. The problem is that if the conceptual space metric is non-Euclidean, then regions may be non-convex. The goal of this section is to describe what the consequences would be if the assumption was of a non-Euclidean metric.

As introduced in Sect. 3.1 above, the formula for the (standard) distance, given a generic Minkowski metric, between two objects (and/or prototypes of concepts) A and B located within an n-dimensional space, is given by the expression:

$$d(A, B) = \left(\sum_{i=1}^{n} \left| X_i^{[A]} - X_i^{[B]} \right|^p \right)^{1/p}$$

where the value of p determines the specific type of distance ($p = 1$, Manhattan; $p = 2$, Euclidean), and it could take any positive real value (not only integer). The boundaries of conceptual regions will then depend on the specific metric chosen, and so will the convex or non-convex character of those regions (as illustrated by the graphs in Fig. 2).

As is evident from Fig. 2, only the (standard) Euclidean metric satisfies the convexity requirement,[19] while the other metrics generate regions that are more or less non-convex.

Consequently, if the metric of conceptual spaces is not Euclidean in a strong sense, then the convexity constraint on regions cannot be mandatory in that very same strong sense; contradicting what Gärdenfors' theory requires of them.

7 Conceptual spaces under a weighted Euclidean metric

Nonetheless, even if the metric of conceptual spaces was Euclidean, it is possible that conceptual regions would not be convex. Obviously, this would not be case, as just explained in the section above, under the standard Minkowski distance which, for the Euclidean case ($p = 2$), is defined as:

$$d(A, B) = \sqrt{\sum_{i=1}^{n} \left(X_i^{[A]} - X_i^{[B]} \right)^2}$$

But let us think for a moment about how, in a theory such as Gärdenfors', concepts are produced. In a first instance, if a particular concept is not innate, then it should

[19] For a formal demonstration of this, see Okabe et al. (1992, p. 57).

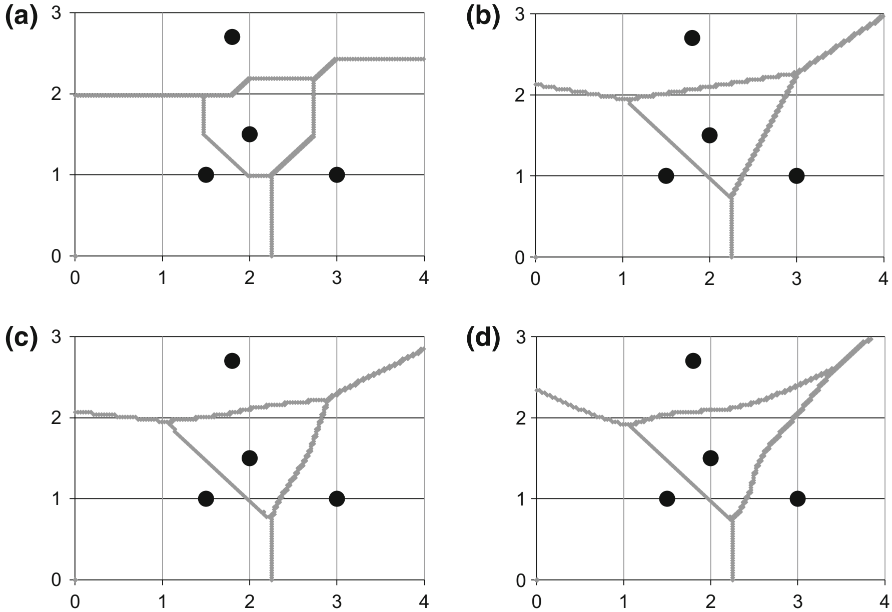

Fig. 2 Boundaries of the conceptual regions resulting from a maximization process implementing the prototype theory of concepts, for four distinct possible metrics. The final prototypes are represented by the *four black dots*, whose coordinates are (1.5, 1), (1.8, 2.7), (2, 1.5) and (3, 1). The boundaries of the conceptual regions are drawn as *dotted dark-grey lines*. **a** Boundaries for the city-block metric (parameter $p = 1$). **b** Boundaries for the Euclidean metric (parameter $p = 2$). **c** Boundaries for a conceptual space that fits the Euclidean metric better than the city-block one (with parameter $p = 1.7$), as happened in Handel and Imai's (1972, p. 110) experiments. **d** Boundaries for a higher-order Minkowski metric (parameter $p = 3$)

have been learnt sometime in the past from a set of particular examples. Additionally, it could be argued that the size of the sample of examples has an effect on how objects are categorized under a particular concept.

For example, let us imagine a subject who had been exposed to hundreds of instances of the concept DOG, but only a few cases of the concept FOX. Then it could be thought that if that same subject were exposed to one new instance of FOX, different from all the foxes already encountered and with a certain resemblance to the concept DOG already acquired, a judgment of the new instance as fitting the concept DOG could be more confidently reached than one of it fitting the concept FOX. That would mean (within a conceptual space theory of the mind), that the subject could ascribe a greater weight to the concept DOG than to the concept FOX (see Fig. 3).

A phenomenon such as this could occur even under a Euclidean metric (that is, even if the underlying conceptual space were Euclidean), where base distances were calculated using the formula above for $d(A, B)$. If objects were categorized as just been described, the distances associated with each concept would be differently weighted depending on the number of examples on which that concept were based. These differently weighted distances would correspond with the non-standard multiplicatively-weighted distances introduced in Sect. 3.1 above. Consequently, the formula for the distance-of-comparison, $d_{Ci}(O, P_{C_i})$, in categorizations of a particu-

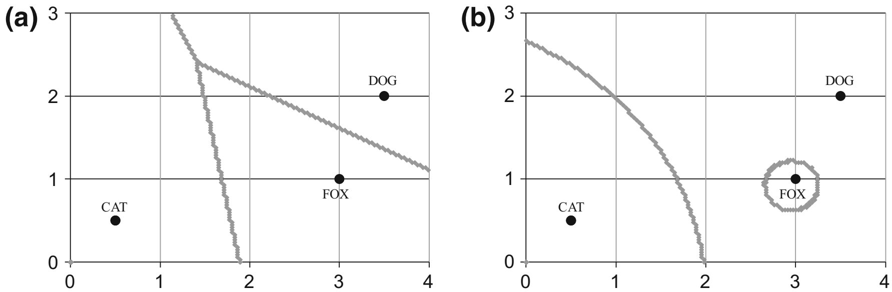

Fig. 3 Boundaries of the conceptual regions resulting from a maximization process implementing the prototype theory, for different weightings of concepts. The three considered concepts are DOG, CAT and FOX, whose prototypes are represented by the *black dots*, with coordinates (3.5, 2), (0.5, 0.5) and (3, 1). The boundaries of the conceptual regions are drawn as *dotted dark-grey lines*. **a** Boundaries for the standard Euclidean space, where the weights of all the prototypes are equal to 1. **b** Boundaries for a non-standard (prototype-weighted) Euclidean space, where the distances-of-comparison for the concepts DOG, CAT and FOX are multiplicatively-weighted by 0.25, 0.4 and 1 respectively (what could happen if the subject had been exposed [i] to a great number of examples of dogs, [ii] to a smaller number of cats, and [iii] only to a very few number of foxes)

lar object O with regard to a given concept C_i (represented by a prototype P_{C_i}) would be:

$$d_{Ci}(O, P_{C_i}) = w_i d(O, P_{C_i})$$

Here, the value of w_i represents the weight associated with each concept, which would be a function of the number of examples, n_i, on which such a concept is based. Indeed, the greater the number of examples, n_i, the greater the relative similarity claimed between O and P_{C_i} (or, equivalently, the lower the distance-of-comparison $d_{Ci}(O, P_{C_i})$), and hence, the lower the weight of the distances w_i. The weight w_i could be, for example, a function ranging from two (if the number of examples is very small) to one (when that number is large enough), as given by $w_i = 1 + 1/n_i$ (see Fig. 4).

The point is that a conceptual space which functioned in this way would produce conceptual regions whose shapes are different from the ones produced by the standard Euclidean metric. Those shapes will be commonly non-convex, what contradicts the assumption with regard to the convexity constraint. The graphs in Fig. 3 contrast the boundaries of convex regions in the standard Euclidean space, with the boundaries of non-convex regions in a prototype-weighted Euclidean space.

Consequently, if concepts were weighted differently (depending on the sizes of their sets of examples) then, even within a Euclidean space, conceptual regions could be non-convex.[20] Of course, the foregoing requires empirical contrast via psychological

[20] For a summary of the properties of a weighted conceptual space, see Okabe et al. (1992, pp. 120–123). One of those properties is that the regions resulting from a multiplicatively weighted Voronoi tessellation do not need to be convex (as shown in Fig. 3), or even connected; and that they can also contain holes. Additionally, according to this kind of approach, the region associated with a particular concept C_i will be convex if and only if the weights of all its adjacent regions are smaller than w_i (in Fig. 3 that is the case of the region associated with the prototype P).

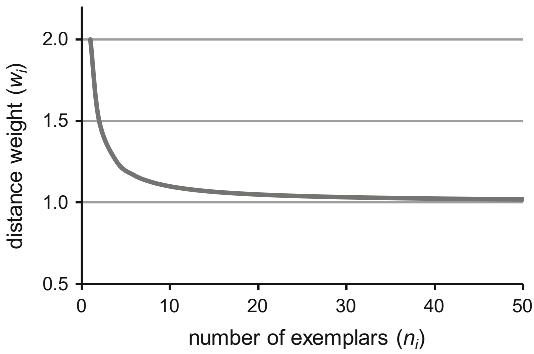

Fig. 4 Representation of the weight function $w_i = 1 + 1/n_i$, that could underlie a non-standard multiplicatively-weighted Euclidean space

research, which will have to decide whether concepts are different weighted or not. Nonetheless, at least from a theoretical perspective, the size of the set of examples from which a certain concept is learnt could influence the reliability of such a concept. On my view, this possibility is significant by itself, beyond the fact that at present there exists or not empirical evidence about it

Therefore, there are important reasons to think that not every concept has the same weight in the conceptual space structure. If this were indeed the case, then those distinct weights would lead to a non-standard Euclidean space, which would result in non-convex conceptual regions.

8 Inner problems of convexity in Gärdenfors' theory

So far I have shown the following. [1] None of the arguments provided by Gärdenfors for the convexity constraint constitutes a compelling reason in favor of that requirement, given that all of them rest on controversial assumptions. [2] His argument for the integral character of conceptual dimensions (in support of a Euclidean metric) is weak; while under a non-Euclidean metric, the structure of regions can be non-convex. [3] Even if the metric were Euclidean the convexity constraint might be not satisfied; if, for example, distinct concepts were differently weighted in terms of the number of examples on which each of them is based.

However it could be the case that, despite all of this, conceptual regions are in fact convex (as assumed by Gärdenfors). In this section, I show that Gärdenfors' convexity constraint is brought into question by his own characterization of conceptual spaces. On the one hand, I will prove that in some cases the regions associated with the properties of a concept are not convex (either taken individually, or as the result of their combination in that concept); while in other cases, the composition of convex regions associated with properties can lead to non-convex concepts (depending on how the properties co-vary over those concepts). On the other hand, I will show that Gärdenfors' definition of properties in terms of convex regions is not compatible with his characterization of verbs as convex regions of vectors from one point to another.

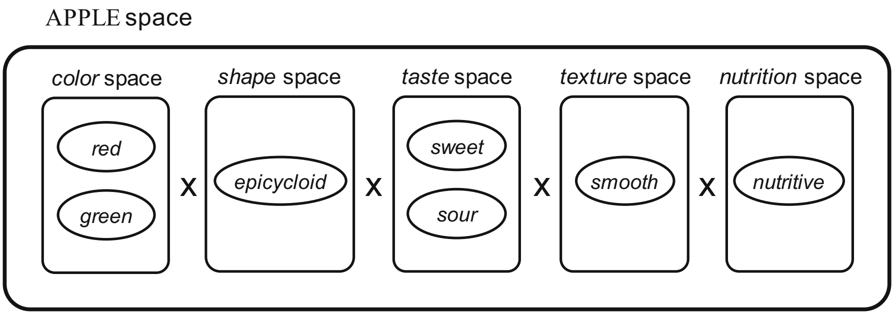

Fig. 5 Inner form of the APPLE conceptual space, as a product space of different quality properties. The APPLE space is represented by the *bigger rounded rectangle*. Properties (such as *red*, *green*, *epicycloid*, etc.) are convex regions represented by the ellipses; for example, the property GREEN corresponds to a convex region of the *color* space, or *color* domain. Quality domains (such as *color*, *shape*, *taste*, etc.) are represented by the *smaller rounded rectangles* (adapted from Fiorini et al. 2014, p. 132)

8.1 On the convexity of properties and concepts

One of the most recent papers co-authored by Gärdenfors (Fiorini et al. 2014) provides a detailed description, absent from previous work, of the inner workings of conceptual spaces. In that paper, Gärdenfors and collaborators represent the inner structure of the APPLE concept by the product space resulting from the properties in those quality domains that form such a conceptual space (as shown in Fig. 5).

Difficulty 1 *Some of the properties are not convex.*
This difficulty could be summed up as follows. There are non-convex physical properties, and it is not easy to conceive a convex approach for the representation of some of those non-convex properties. The first point is largely uncontroversial, given that the physical shape of many objects is not convex, as happens with the shape of an apple.

With regard to the shape properties, Gärdenfors proposes different models for representing them, suitable for different kinds of shapes. Nevertheless, none of them is proper for the characterization of general shapes and, in particular, for a convex representation of the shape of an apple, as it will be shown in the following points:

(A) The approach followed to represent rectangles (Gärdenfors 2000, pp. 93–94; 2014, pp. 35–36) by the conditions satisfied by their quadruples of points in R^2 can only be applied to very basic geometrical shapes (not including the epicycloid).
(B) The model proposed for the analysis of general shapes (Gärdenfors 2000, pp. 95–96; 2014, pp. 121–122), based on the work of Marr and Nishihara (1978), could be more or less applicable to the case of the shapes of animals, as a combination of cylinders (associated with their different parts) together with information about how those cylinders are joined, but not to the shapes of arbitrary objects.

- Therefore, neither of these two models allows us to represent the shape of an apple, distinguishing it from the shape of a lemon, pear or melon. And, although the second is useful for characterizing movements and actions, neither of them is compelling as a model for general shapes.

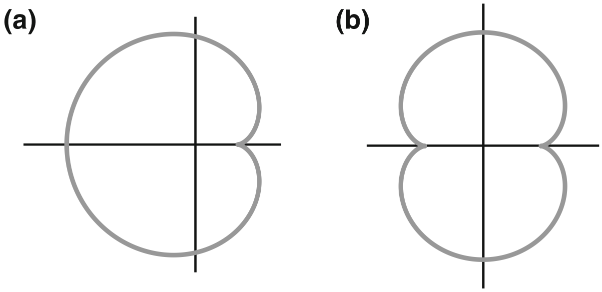

Fig. 6 Epicycloid curves representing the ideal two-dimensional (2D) contour of an apple. The apple's ideal shape will be the three-dimensional (3D) surface resulting from the rotation of any of these curves around the *horizontal axis*. **a** Epicycloid curve with ratio $n = 1$. **b** Epicycloid curve with ratio $n = 2$

(C) Thirdly, his approach to locative prepositions (Gärdenfors 2014, pp. 205–214) leads to an accurate formalization of the meaning of *near*, *far*, *inside*, *outside*, *beside*, etc. in terms of a polar coordinate system. In light of this, it seems that Gärdenfors' aim is to transfer how these prepositions are applied to shapes in the physical world, to the shapes of their associated conceptual spaces.[21] And the same can be said regarding to his description of the meaning of *bumpy*, as a structure in physical space constituted by "an even (but continuous) distribution of values on the vertical dimension of a horizontally extended object" (Gärdenfors 2014, p. 246).

However, a direct translation of shapes from the physical space to a convex representation within a conceptual hyperspace is only possible if the shape of the considered object is convex. The problem is that the shape of many objects is not convex, as happens with the EPICYCLOID[22] for the case of apples. An epicycloid is a plane curve generated by the path of a point on a smaller circle (with radius r) as that circle rolls around a larger fixed circle (with radius nr, where n is an integer). The epicycloid is given by the following parametric equations:

$$x(\theta) = r(n+1)\cos\theta - r\cos[(n+1)\theta]$$
$$y(\theta) = r(n+1)\sin\theta - r\sin[(n+1)\theta]$$

An apple shape could be associated with an epicycloid with a value of n equal to 1 or 2 (see Fig. 6).

Nonetheless, neither of these 2D curves (and consequently, none of their associated 3D surfaces) is convex, because in both cases it is possible to find pairs of points within the regions they bound such that some points between them do not belong to the same region.

[21] The problem is that, for the convexity constraint to be met by the regions characterizing these prepositions, the convexity of the objects to which they apply is necessary.

[22] Although Fiorini, Gärdenfors and Abel describe the apple's shape as a cycloid, in fact the shape corresponds to an epicycloid.

Finally, this problem extends from the apple example to many other object categories whose shapes are not convex. And, although the fact that Gärdenfors is not able to provide a method for the convex representation of non-convex shapes (such as the shape of an apple) does not constitute a proof that no method exists, it is an evidence for the difficulty of conceiving a natural way of representing a non-convex shape by means of a convex conceptual region.

Difficulty 2 *The conceptual region resulting from the combination of convex properties (belonging to the same domain) can be non-convex.*
This characterization of the APPLE space sheds a great deal of light on how conceptual spaces are supposed to work internally, especially regarding the following point: different properties in the same domain can be associated with the same concept (for example, the properties RED and GREEN in the *color* domain, or SWEET and SOUR in the *taste* domain, for the case of the APPLE concept).[23]

Here the problem is that two properties from the same domain (RED and GREEN, for instance) cannot be composed into a product space. Let us recall here that the product space, R, of a set of constitutive properties Q_1, Q_2, ..., Q_n, is equal to the set of objects[24] belonging simultaneously to Q_1, Q_2, ..., and Q_n. For instance, if the APPLE space were constituted only by the *shape* and *texture* spaces previously shown, then a particular object would be an apple if it were EPICYCLOID and SMOOTH. Or, from a logical point of view, for an object to be categorized as an apple, it is necessary (but not sufficient) that the following conjunction of properties is satisfied over those quality domains:

[23] In this case it could be argued that, because it is evident that the disjunctive combination of convex properties in the same domain can fail to be convex, Gärdenfors could not have been unaware of it. This leads to the question of to what extent Gärdenfors is committed to convexity. Here I will try to show that Gärdenfors is strongly committed to the convexity of concepts, but before a terminological clarification is needed.
In Gärdenfors (2000) he distinguished between properties (defined as convex regions) and concepts (defined as sets of convex regions), but it is not possible to find there an explicit assertion about the convexity or non-convexity of concepts. However, things change in his latter works, where we find statements like "as proposed in Gärdenfors (2000), concepts can be modeled as convex regions of a conceptual space" (Warglien and Gärdenfors 2013, p. 2171), or "the convexity of concepts is also crucial for ensuring the *effectiveness of communication*" (Gärdenfors 2014, p. 26). Nonetheless, it is not clear that in these last quotes Gärdenfors is referring with the term "concept" the same as in Gärdenfors (2000), because in Gärdenfors (2014) the notion "object category" began to play the role played by the notion "concept" in Gärdenfors (2000). Withal, in Gärdenfors (2014), he does not explicitly assert that object categories are represented by convex regions. Notwithstanding this, in these recent works (Warglien and Gärdenfors 2013; Gärdenfors 2014) he tries to explain human communication via a "meeting of minds", using a fix-point argument. Such an argument requires the convexity of concepts: "the concepts in the minds of communicating individuals are modeled as convex regions in conceptual spaces (...) If concepts are convex, it will in general be possible for interactors to agree on joint meaning even if they start out from different representational spaces." (Warglien and Gärdenfors 2013, p. 2165). But, this argument is expected to apply at least to noun phrases (ib., p. 2170) and, thus, to object categories (as the APPLE concept), which should also be represented by convex regions. On my view, this proves that Gärdenfors is firmly committed to the convexity of both properties and concepts.

[24] It may be thought that the conceptual region R was equal to *the* (not *a*) set of objects belonging simultaneously to Q_1, Q_2, ..., and Q_n. However, that is not the approach followed by Gärdenfors (2014, p. 29), who accepts the possibility of a non-rectangular conceptual space, which does not contain the whole set of points belonging simultaneously to all its constitutive properties (as is the case of the graphs shown in Fig. 8). That is the reason why the following logical conditions are necessary, but not sufficient.

$$(shape = \text{EPICYCLOID}) \wedge (texture = \text{SMOOTH})$$

All this works if the properties considered belong to different domains. The problem is that when two (or more) properties belong to the very same domain, they cannot be composed into a product space, because in this case the product space would not include the desired set of objects. For instance, if we now included the *color* domain for the case of the APPLE concept, the resulting product space would have to satisfy the condition:

$$(color_1 = \text{RED}) \wedge (color_2 = \text{GREEN}) \wedge (shape = \text{EPICYCLOID}) \wedge (texture = \text{SMOOTH})$$

This condition would certainly include all the red-and-green apples,[25] but not the green (but not red) apples, nor the red (but not green) apples. This is so because the kind of composition required when two or more properties belong to the very same domain is not their product space, but their addition space, that is, the region resulting from the union of the regions associated with those properties. This kind of composition could be identified with a logical disjunction, so the condition associated with the APPLE space could be better expressed (with a unique *color* dimension) as:

$$[(color = \text{RED}) \vee (color = \text{GREEN})] \wedge (shape = \text{EPICYCLOID}) \wedge (texture = \text{SMOOTH})$$

The problem is that the conceptual space resulting from the addition of the RED and GREEN properties is not convex, given that the ORANGE color establishes a discontinuity between them; as is obvious from their representation in the color spindle (Fig. 7). Consequently, the resulting *color* space (associated with the APPLE concept) is not convex.

An even clearer case would be that associated with the SWAN conceptual space, which would be constituted (following Gärdenfors' approach) by the product space resulting from a set of properties in the quality domains *color*, *shape*, etc. In this case, two different properties (BLACK and WHITE) are represented in the *color* domain. Those two properties should be represented by convex regions (in fact, points) within the *color* domain, but there is no path within the color spindle between them that only passes through points representing the colors of a swan. In this case, the combination of the BLACK and WHITE properties determines a disconnected region, so it cannot be convex (or even star-shaped).[26]

[25] At the expense of considering two different color dimensions ($color_1$ and $color_2$) as constitutive of the APPLE conceptual space, given that if both color dimensions were the same, the set of objects satisfying this condition would be void. Here I will not enter into the discussion of the problems associated with such implications.

[26] Therefore, the SWAN conceptual space is a problem for any theory which attributes a mandatory character to the connectedness requirement for the geometry of conceptual regions. Obviously, this applies to any criterion stronger than the connectedness one (as is the case with the star-shapedness and convexity requirements).

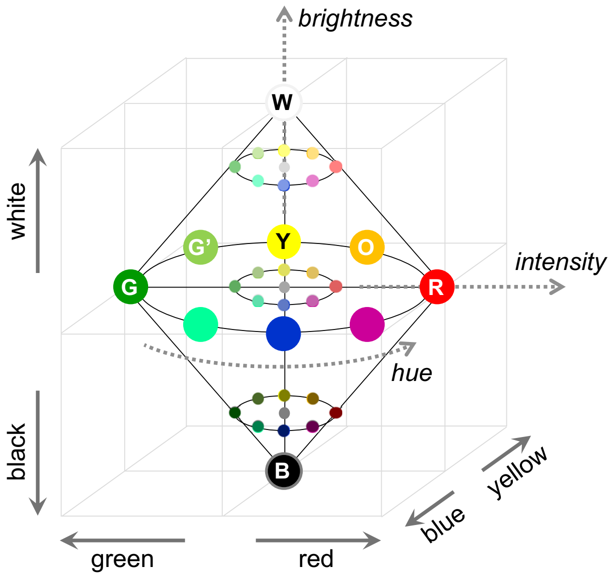

Fig. 7 Representation of the color spindle and its constitutive dimensions: *hue*, *intensity* and *brightness*. The relevant colors for the examples provided are denoted by their initials: R = red, G/G′ = green, O = orange, Y = yellow, W = white, B = black (adapted from Churchland 2005, p. 536). (Color figure online)

Difficulty 3 *The space resulting from the covariation of convex regions can be non-convex.*
Even if a concept C is constituted by properties represented by convex regions (as Gärdenfors assumes), depending on how the properties of the instances (or particular cases) of C co-vary, the conceptual region R associated to C might be convex or not.

Let us now consider a concept C composed by two properties F_1 and F_2. If F_1 and F_2 were properties located in domains constituted by only one dimension (d_1 and d_2, respectively), then the region R representing C will be situated in a conceptual space composed by those two dimensions d_1 and d_2. For instance, this is the case of the MOUNTAIN concept, whose properties (*width* and *height*) are represented by one-dimensional domains (Adams and Raubal 2009, p. 258; quoted in Gärdenfors 2014, p. 29).

Let us assume further that those properties F_1 and F_2 are represented by the regions (more specifically, intervals) Q_1 and Q_2 within the dimensions d_1 and d_2, respectively. For the case of the MOUNTAIN concept, if d_1 and d_2 represent the *width* and *height* of the mountain, respectively, those regions could be given by the intervals $Q_1 = (1500, 13000)$ and $Q_2 = (1200, 8000)$ (all of them expressed in meters).

However, it could happen that C were represented by a region R which did not cover completely the Cartesian product $Q_1 \times Q_2$ of its constituent properties. In that case, there will exist pairs of points (q_1, q_2) which belong to $Q_1 \times Q_2$, but do not belong to R, that is, $\exists(q_1, q_2)((q_1 \in Q_1) \wedge (q_2 \in Q_2) \wedge ((q_1, q_2) \notin R))$. For instance, the MOUNTAIN's conceptual region might not be rectangular, but triangular (see Fig. 8a):

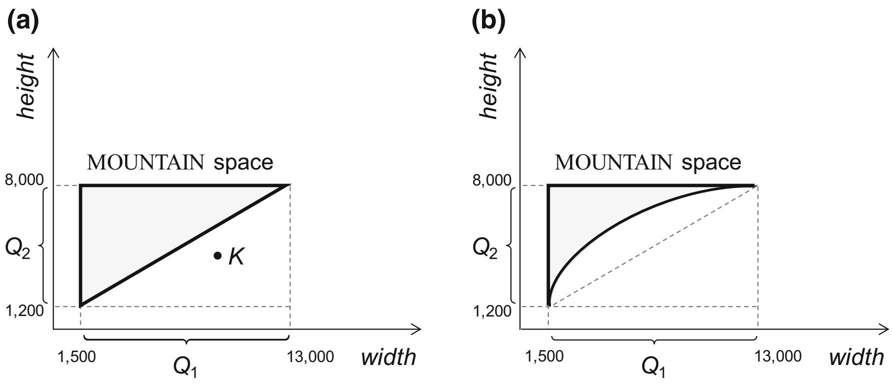

Fig. 8 Two different ways in which the covariation between the *width* and *height* dimensions (constitutive of the MOUNTAIN concept) can happen. **a** Covariation resulting in a triangular conceptual region. **b** Covariation resulting in a triangle with a concave curve as the hypotenuse

> If a formation is very high, its width will not matter much; it will still be a mountain. However, a lower and very wide formation might not be called a mountain. Thus, the region in the product space that represents mountain has more or less a triangular shape. (Gärdenfors 2014, p. 29)

In such a case, an upward projection of the earth's surface whose width and height were given by the pair (9000, 4000) (represented as K in Fig. 8a), would belong to $Q_1 \times Q_2$, but would not be called a mountain.

Withal, if the conceptual space associated with MOUNTAIN is triangular, the convexity of the MOUNTAIN's conceptual region is guaranteed.

Nonetheless, if the region R (representing C) did not cover completely the Cartesian product $Q_1 \times Q_2$ of the constituent properties of C, it could happen that R is not even convex. For instance, in Adams and Raubal's example the hypotenuse of the triangle which delimits the conceptual region of MOUNTAIN may not be a straight line, but a concave curve (as shown in Fig. 8b). Therefore, depending on how the properties of a concept C co-vary, its conceptual region will be convex or not.

8.2 On the compatibility of the convexity of properties and verbs

Lastly, when Gärdenfors extends his conceptual space theory to the semantics of verbs, such an extended framework introduces a general problem (which could be described as structural). It is associated with his definition of verbs, and closely related to his basic conception of property. In this case, the problem is that his characterization of *verb* meanings, as vectors from one point to another, is not compatible with a definition of properties and concepts in terms of convex regions.

In his extended theory, Gärdenfors identifies *states* and *changes* with zero and non-zero vectors; and based on them he defines *events* as changes in the state of a patient (usually due to the action of an agent). The problem is that, strictly speaking, *states* and *changes* cannot be identified with points and single vectors, respectively, if *properties* are not represented by points, but by (allegedly) convex regions. In virtue of this, *states* should be represented by regions; and *changes* therefore ought to be represented by sets of vectors from every point in the region associated with the initial property, to every point in the region associated with the final property.[27]

The same can be said with regard to the result vectors associated with a given verb. Gärdenfors defines a *verb* as a change in the properties of an object; that is, as a movement in its representation within the conceptual space. Based on this, such a change is represented by means of a vector from the position of the initial object to that of the final object. However, and given that a state is, in fact, not represented by a point (but by the region associated with the property described by such a state), a verb cannot be represented merely as a vector (or a mapping from one point to another), but must be represented as a mapping from one region to another.

Thus, a *verb* should be represented not by a vector (or convex set of paths, with only one origin and one endpoint), but by a set whose elements are convex sets of paths (each with a different origin and/or end).

Obviously, here I have not proved that those sets of vectors which should represent verbs cannot be convex. To my knowledge, it is hard to see in which sense it may be said that a set C constituted by all the pairs of points (x, y) such that, $x \in X$ and $y \in Y$, where X and Y are convex sets, is convex. Notwithstanding, the burden of proof lies on the side of Gärdenfors. If he wants (i) to provide a unifying framework for the semantics of verbs, nouns and adjectives (Warglien et al. 2012), and (ii) to explain the semantics of verbs by changes of states (ib.), he has to prove that a set C as the one just described is convex.

9 Conclusions

One of the main theses of Gärdenfors' (2000, 2014) conceptual space theory is the convexity constraint on the geometry of the conceptual regions associated with properties, concepts (or object categories), actions, verbs, prepositions and adverbs. Nonetheless, in this paper I have shown that such a convexity constraint is problematic; both from a theoretical perspective, and with regard to the inner workings of the theory itself.

[27] This is the same notion of change that Gärdenfors has in mind when he says the following:

> In general, a change of state is not represented by a specific vector. Instead, it can be represented by a category of changes of state. (…) If the start point is set as the origin, one can represent a category of change events as a region of end points. (…) For example, going "upwards" in a two-dimensional space will correspond to a convex region of points located in a cone to the "north" of the origin. (Warglien et al. 2012, p. 162)

However, this is not the only possible way to generalize the notion of change. Another possibility would be that the knowing subject S knew neither the initial point nor the final point of the change, and that the initial point could not be set as the origin. This would happen if John said to S, "the leaves were yellowed by disease", but S does not know the tree whose leaves were referred to. In this case S knows neither the exact start point (a kind of green) nor the exact end point (a kind of yellow) of the change expressed by "yellowed", so he will have to represent the properties associated to the initial and final states by regions (the ones associated to the GREEN and YELLOW colors). Consequently, the change expressed by "yellowed" will be represented as the set of vectors going from every point in the region representing GREEN to every point in the region representing YELLOW.

On the one hand, I have shown that none of Gärdenfors' arguments in favor of the convexity requirement compels us to accept it as a mandatory criterion for the geometry of regions. [1] With regard to his first argument, everything that can be said concerning the co-implication of the prototype theory and the convexity of regions, could also be said regarding the star-shapedness constraint on regions. [2] In relation to the cognitive economy argument, it depends on the controversial assumption that handling convex regions requires fewer computational resources than handling regions with capricious forms. [3] Regarding the perceptual foundation argument, it relied on the hypothesis that perceptual and conceptual domains share the same geometrical structure, which might not be the case. [4] Finally, his argument concerning a more effective communication is based on studies that assume the standard Euclidean metric, but such an assumption is far from trivial.

On the other hand, and with regard to the kind of metric underlying conceptual spaces, under the standard Euclidean metric assumed by Gärdenfors (that is, Euclidean and non-weighted distances), the convexity of regions would indeed be guaranteed. However, the question regarding the type of metric that can underlie conceptual spaces is an empirical one; and all of the evidence provided by Gärdenfors in support of the standard Euclidean metric is controversial. Firstly, the Euclidean metric cannot be based on the integral character of domains, requiring empirical evidence independent from the one associated to the non-separability of dimensions. Secondly, the empirical evidence referred to in favor of the integral character of domains (and, in consequence, in favor of the Euclidean metric and the convexity of regions) comes from a very small number of perceptual domains; things might not work in exactly the same way in other perceptual and in non-perceptual domains. Thirdly, none of the work cited identifies integral domains with the Euclidean metric perfectly, but rather with a metric that is more similar to the Euclidean than to the city-block; and such a kind of metric does not lead to convex conceptual regions. Due to all of this, if the metric underlying conceptual spaces were standard, it may be that it would not be Euclidean in a strong sense; and in that case, it has been shown that the convexity constraint on regions is not valid.

Additionally, it has been proved that, even if the metric underlying conceptual spaces were Euclidean, regions could be non-convex if the distances-of-comparison in categorizations were differently weighted; depending, for example, on the number of examples on which each concept is based. That is, convexity is guaranteed only under the standard Euclidean metric: not under a weighted Euclidean metric. The problem is that, even if the psychological space is Euclidean, there are good reasons in favor of a non-standard multiplicatively-weighted determination of distances; under which, conceptual regions could be non-convex.

Finally, even if none of the above problematic possibilities were the case, Gärdenfors' convexity constraint is brought into question by his own characterization of the inner workings of conceptual spaces. The problems could be summed up as follows. [I] Some of the allegedly convex properties of concepts are not convex, as happens with those associated with the *shape* domain, and it is not clear how they could be represented in a convex way. [II] The conceptual region resulting from the combination of two (or more) convex properties belonging to the same domain can be non-convex, and the same happens for its associated concept. [III] The space resulting from the covariation of different convex regions could be non-convex; as a theoretical possibility that needs for more empirical research. [IV] Gärdenfors definition of verbs, as vectors from one point to another, is not compatible with a definition of properties and concepts in terms of convex regions.

Based on all this, I conclude that the mandatory character of the convexity requirement for regions in any similarity space theory of concepts (and so in Gärdenfors' conceptual spaces) should be reconsidered, in favor of a weaker constraint, such as a non-obligatory version of the star-shapedness constraint. Notwithstanding, this paper should not be seen as a defense of a mandatory star-shapedness requirement, given that such a constraint has its own problems, most of which are not mentioned here.[28]

Acknowledgements This research was carried out under a PhD scholarship from the University of the Basque Country, and financially supported by the Spanish Ministry of Economy and Competitiveness research projects FFI2011-30074-C02-02 and FFI2014-52196-P. I wish to give special thanks to Agustin Vicente for his continued advice and encouragement. A partial version of this paper was presented at the SPE8 Colloquium (2015). I am grateful to the audience there for valuable comments. Finally, I am also thankful to Toffa Evans for his assistance with the manuscript.

Funding This study was financially supported by the Spanish Ministry of Economy and Competitiveness (research projects FFI2011-30074-C02-02 and FFI2014-52196-P), and was carried out under a PhD scholarship from the University of the Basque Country.

Compliance with ethical standards

Conflicts of Interest The author declares that he has no conflict of interest.

References

Adams, B., & Raubal, M. (2009). Conceptual space markup language (CSML): Towards the cognitive semantic web. In *Proceedings of the third IEEE international conference on semantic computing (ICSC 2009)* (pp. 253–260). Berkeley, CA: IEEE Computer Society.
Churchland, P. (2005). Chimerical colors: Some phenomenological predictions from cognitive neuroscience. *Philosophical Psychology*, *18*(5), 527–560.
Fairbanks, G., & Grubb, P. (1961). A psychophysical investigation of vowel formants. *Journal of Speech and Hearing Research*, *4*(3), 203–219.
Fiorini, S. R., Gärdenfors, P., & Abel, M. (2014). Representing part–whole relations in conceptual spaces. *Cognitive Processing*, *15*(2), 127–142.
Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.
Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. Cambridge, MA: MIT Press.
Gärdenfors, P., & Warglien, M. (2012). Using conceptual spaces to model actions and events. *Journal of Semantics*, *29*(4), 487–519.
Gärdenfors, P., & Warglien, M. (2013). Semantics, conceptual spaces, and the meeting of minds. *Synthese*, *190*(12), 2165–2193.
Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
Gauker, C. (2007). A critique of the similarity space theory of concepts. *Mind & Language*, *22*(4), 317–345.
Goldstone, R. L., & Son, J. Y. (2005). Similarity. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 13–36). Cambridge: Cambridge University Press.

[28] Nonetheless, these problems would also be associated with the convexity constraint (given that every convex region is also a star-shaped region).

Handel, S., & Imai, S. (1972). The free classification of analyzable and unanalyzable stimuli. *Perception and Psychophysics*, *12*(1), 108–112.

Jäger, G. (2007). The evolution of convex categories. *Linguistic and Philosophy*, *30*(5), 551–564.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Lee, D. T. (1980). Two-dimensional Voronoi diagrams in the L_p-metric. *Journal of the Association for Computing Machinery*, *27*(4), 604–618.

Maddox, W. T. (1992). Perceptual and decisional separability. In G. F. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147–180). Hillsdale, NJ: Lawrence Erlbaum.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional shapes. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, *200*(1140), 269–294.

Melara, R. D. (1992). The concept of perceptual similarity: From psychophysics to cognitive psychology. In D. Algom (Ed.), *Psychophysical approaches to cognition* (pp. 303–388). Amsterdam: Elsevier.

Okabe, A., Boots, B., Sugihara, K., & Chiu, S. N. (1992). *Spatial tessellations: Concepts and applications of Voronoi diagrams*. New York: Wiley.

Quine, W. V. (1969). Natural kinds. *Ontological relativity and other essays* (pp. 114–138). New York: Columbia University Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum Associates.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Sivik, L., & Taft, C. (1994). Color naming: A mapping in the NCS of common color terms. *Scandinavian Journal of Psychology*, *35*(2), 144–164.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Warglien, M., & Gärdenfors, P. (2013). Semantics, conceptual spaces, and the meeting of minds. *Synthese*, *190*(12), 2165–2193.

Warglien, M., Gärdenfors, P., & Westera, M. (2012). Event structure, conceptual spaces and the semantics of verbs. *Theoretical Linguistics*, *38*(3–4), 159–193.