# Convexity is a testable prediction in the theory of conceptual spaces: Reply to Hernández-Conde

Peter Gärdenfors[1]

*Abstract*

This article is a rejoinder to Hernández-Conde's (2016) criticism of the convexity criterion in the theory of conceptual spaces. His arguments in general claim that the convexity criterion could be false and that it therefore is problematic for the theory. However, this is a misunderstanding since the convexity criterion is put forward as an empirically testable thesis.

José Hernández-Conde (2016) criticizes the convexity requirement in the theory of conceptual spaces. His article presents a long list of cases where he argues that the convexity criterion does not have a ”mandatory character”. It contains some valid points, to which I return below. However, the article is based on a basic misunderstanding of the role of the convexity criterion and it contains some direct errors.

Most of the arguments by Hernández-Conde have the following character:

(1) If convexity is a valid criterion, then X would follow.

(2) It *could be* that X is not true.

(3) Hence, the convexity criterion is not supported.

This kind argument presupposes, however, that convexity is a necessary condition for applications of conceptual spaces (Hernández-Conde repeatedly writes about the “mandatory character” of the convexity criterion). I believe that the structure of the argument reveals a basic misunderstanding. The convexity criterion is not proposed as something that necessarily holds of an application, but as a *testable prediction* (Gärdenfors 2000, 2014). The convexity criterion is what furnishes the theory of conceptual spaces with most of its empirical content.

The ‘could be’ argument is repeated throughout the article. I will not discuss all cases, but focus on a few. A typical example is section 4.3 where he argues that the convexity of the regions of the color domain is “no guarantee” that the convexity criterion will work in non-perceptual domains. True, but the criterion is put forward as a testable hypothesis for other domains. If the criterion would be guaranteed to work, it would have no empirical content.

In section 3.2, Hernández-Conde writes that I am “strongly committed to the thesis that the shapes and boundaries of conceptual regions are produced by a Voronoi tessellation of the conceptual hyperspace” (this Hernández-Conde calls Thesis V). Hernández-Conde notes that I never express this thesis openly, but he claims that I accept it tacitly. That is not true – for several reasons. Firstly, for many basic dimensions – such as length, weight, age, temperature, time – there exist no prototypes and, consequently, Voronoi

*Mail to: Peter.Gardenfors@lucs.lu.se*

[1] Lund University Cognitive Science, Department of Philosophy, LUX, Box 192, S-22100 Lund, Sweden.

tessellations are not definable for these dimensions. However, adjectives that are used to express properties determined by such dimensions ('long' vs. 'short', 'heavy' vs. 'light', etc) nevertheless represent convex regions. Similarly, at the beginning of section 4.1 Hernández-Conde writes that "concepts show prototypical effects". Yes many do, but not the examples above. There are connections between conceptual spaces and prototype theory, but there is no "mutual dependence". For example, Hernández-Conde argues correctly that "[t]he only things that should be expected by a consistent prototype theorist is the star-shapedness of conceptual regions." That's true, but I put forward the stronger convexity criterion as an empirical hypothesis. This is one aspect (among several) where conceptual spaces lead to richer predictions than prototype theory. Secondly, in the case when a domain has a Euclidean metric, there are convex tessellations of the domain that are not generated by the Voronoi mechanism, so in this subcase, Thesis V is stronger than the convexity criterion. Thirdly, when a domain has a metric that is not Euclidean, Voronoi tessellations may lead to non-convex regions, so in this case Thesis V and the convexity criterion are in conflict. So if I accept the convexity criterion, I cannot endorse the full Thesis V as the same time. However, Hernández-Conde writes as if I do. Finally, when he in section 4.4 accuses me of *petitio principii*, he also assumes that I take Thesis V for granted.

Prototype theory in itself does not say anything about a possible geometric structure of the representations of concepts. So Hernández-Conde's claim in section 4.2 that "those who adopt the prototype theory of concepts should expect a representation of concepts and properties as convex regions" is misleading. My more specific position should rather be formulated as: "Given a conceptual space and given prototypes of concepts as locations in the space, the convexity of regions can be explained by the mechanism of Voronoi tessellation." However, there might be other explanations. Which is the best one is an empirical question.

In sections 5-7, Hernández-Conde argues that if the metric of a space is not Euclidean, then the tessellation into regions generated by a Voronoi classification *could* result in non-convex regions. True, but this is again an empirical question whether concepts that depend on psychological spaces with a non-Euclidean metric actually correspond to non-convex regions.

At the beginning of section 5, Hernández-Conde writes that "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 that a city-block metric". This is a misrepresentation, since the fact that a Euclidean metric fits the empirical data better that a city-block metric is used in the psychological literatures as one criterion (among others) for *deciding* whether dimensions are integral or separable (see Gärdenfors 2000, pp. 24-26 for a presentation of some criteria and Johannesson 2002 for an extensive discussion). This criterion is thus a part of a definition and not an argument for the Euclidean metric.

In section 7, Hernández-Conde proposes another distance measure that combines prototypes and the number of examples on which a concept is based. Hernández-Conde's point is to show that with such a distance measure, the resulting regions would not necessarily be convex. This is true, but, again, it is an empirical question which rule for categorization fits the data best. As far as I know, there is no psychological data that supports the categorization method he suggests. My hypothesis is still that convex partitionings have more empirical support.

Another misunderstanding occurs in Hernández-Conde's discussion in Section 8.1 of how shapes are represented in conceptual space (in relation to the apple example). I have put forward the prediction that shapes correspond to convex regions of the shape domain. Admittedly, this prediction is not directly testable unless I specify the structure of the shape domain (I discuss some possible approaches in Gärdenfors 2000, Section 3.10.2 and Gärdenfors 2014, Section 6.3). However Hernández-Conde argues that this prediction is false, since the epicycloid shape of apples is clearly non-convex. This is, however, a clear misunderstanding of how shapes are represented in conceptual spaces. The claim is not that the shape of an object is convex in physical space, but that the *class of shapes* of a category forms a convex region in the shape domain. Thus if an apple A has a particular shape and apple B another shape then any shape *between* that of A and B would also be a example of an apple shape. If apple shapes are (approximations of) epicycloids as generated by the pair equations provided by Hernández-Conde for a range of radiuses r, then this class is trivially convex in the sense that r is the only variable, so that if the shape of A is determined by $r_1$ and that of B by $r_2$, then any value between $r_1$ and $r_2$ is also generates an epicycloid belonging to the class.

In section 8.1, Hernández-Conde also argues that the concept of a swan is a counterexample to my definition of object categories since swans are either black or white, so the convexity criterion is violated. However, color is not one of the determining properties for 'swan'. Apart from property regions that are part of all categories falling under the superordinate category 'bird', the shape domain is presumably the most characteristic. This means that the representation of the category of swans does not exclude any color. If a yellow bird that in all other respects had the properties of a swan was discovered, it would be categorized as a swan (this is what happened when the black swans in Australia were observed by the European explorers). In brief, this is not a counterexample, but a fairly standard case of object categorization mechanism.

In conclusion, I am grateful to Hernández-Conde for pointing out so many cases where the convexity principle *could* be violated. This shows that the principle is rich in empirical content. The cases that Hernández-Conde discusses can, and should be, empirically tested. It pleases me that Hernández-Conde does not present any valid counterexamples to the prediction, but only points out they *could* turn out to be false. The upshot is that, in the absence of counterexamples, the features that Hernández-Conde views as problems are instead *strengths* for the theory of conceptual spaces.

## References

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.

Hernández-Conde, J. V. (2016). A case against convexity in conceptual spaces, *Synthese*, DOI 10.1007/s11229-016-1123-z.

Johannesson, M. (2002). *Geometric Models of Similarity*. Lund: Lund University Cognitive Studies 90.