Nonmonotonic Reasoning, Expectations Orderings, and Conceptual Spaces

Matías Osta-Vélez^{1,2} · Peter Gärdenfors³

Accepted: 25 October 2021 / Published online: 21 November 2021
© The Author(s) 2021

Abstract

In Gärdenfors and Makinson (Artif Intell 65(2):197–245, 1994) and Gärdenfors (Knowledge representation and reasoning under uncertainty, Springer-Verlag, 1992) it was shown that it is possible to model nonmonotonic inference using a classical consequence relation plus an expectation-based ordering of formulas. In this article, we argue that this framework can be significantly enriched by adopting a conceptual spaces-based analysis of the role of expectations in reasoning. In particular, we show that this can solve various epistemological issues that surround nonmonotonic and default logics. We propose some formal criteria for constructing and updating expectation orderings based on conceptual spaces, and we explain how to apply them to nonmonotonic reasoning about objects and properties.

Keywords Nonmonotonic inference · Typicality · Expectations · Conceptual spaces

1 Introduction

Logic has been defined as the science of correct reasoning. Classical logic, however, builds on two assumptions that limit the scope of reasoning: Firstly, that inference is a relation between sentences (or propositions). Secondly, that argument validity depends exclusively on the formal structure of the premises and the conclusion, and is independent of their meaning or the context in which the inference is drawn (Gärdenfors 1992). Furthermore, deductive logic is *informationally conservative* in the sense that the information in the conclusion is implicit in the premises. Thus, in deductive reasoning, drawing a conclusion from a set of premises does not require the agent to

✉ Matías Osta-Vélez
matiasosta@gmail.com

¹ Center for Mind and Cognition, Institute for Philosophy II, Ruhr-University Bochum, Bochum, Germany

² Instituto de Filosofía, FHCE, Universidad de la República, Montevideo, Uruguay

³ Department of Philosophy and Cognitive Science, Lund University, Lund, Sweden

use semantic knowledge about the premises but just to exploit syntactic properties of logical form.

Everyday reasoning, however, builds on more than the logical form of explicit premises. Partial information and uncertainty are pervasive and, as a consequence, our inferential mechanisms can hardly afford to be informationally conservative (Oaksford and Chater 2009). Instead, for making sense of our environment, we need to constantly take risks and use our background knowledge in productive ways in order to complement the information that is explicit in the premises.

One particular way in which this use of background knowledge expresses itself is through our expectations about the world.¹ For instance, if we know that Maria is from Spain, we expect her to have dark hair and to speak Spanish; or if we are driving a car and we spot a person waiting on one side of the road, we expect her to intend to cross it. In general, our expectations about the world are crucial for guiding our reasoning and action in everyday life, and they build directly on the structure of our background knowledge.

In earlier work (Gärdenfors 1992; Gärdenfors and Makinson 1994), it has been argued that much of nonmonotonic logic can be reduced to classical logic with the aid of an analysis of the expectations working as hidden premises in arguments. The guiding idea is that when we try to find out whether a conclusion *C* follows from a set of premises *P*, the background information that we use for inferring does not only contain the premises in *P* but also information about what we expect in the given situation, so that we end up with a larger set of assumptions.

Such expectations can be expressed as default assumptions, that is, statements about what is normal or typical. They include not only our knowledge as limiting case but also other beliefs that are regarded as plausible enough to be used as a basis for inference as long as they do not give rise to inconsistencies. Thus expectations are defeasible in the sense that if the premises in *P* conflict with some of the expectations, we do not use them when determining whether *C* follows from *P*. Expectations are used basically in the same way as explicit premises in logical arguments; the difference being that expectations are, in general, ‘more defeasible’ than the premises.

Expectations function as hidden assumptions. However, when evaluating their role in arguments it is important to note that they do not all have the same strength. For example, in certain cases we consider the relation among two propositions to be strong enough to work as a rule that is almost universally valid, so that an exception to it would be extremely unexpected; while in other situations this relation could be better described as rules of thumb that we use if we need more precise information. For instance, while walking on the sidewalk, we expect the ground to be solid enough to support our body weight; but when we are hiking in the snow this expectation will be weaker and therefore we will walk carefully in order to avoid sinking. An exception to the latter type of rule is not unexpected to the same degree as in the former case. In brief, our expectations are all defeasible, but they exhibit varying degrees of defeasibility.

¹ It should be noted that our choice of the word ‘expectation’ should not be confused with the notion of ‘expected utility’ in decision theory. ‘Expected utility’ has to do with expectations of the values of various outcomes, while our notion of expectation concerns beliefs about the world. Our use of ‘expectation’ thus comes closer to the everyday use. Furthermore, we do not assume that expectations should be modelled in terms of probabilities.

In this article, our main objects of study are expectation orderings as a way of summarizing degrees of defeasibility. An argument in favor of using expectation orderings for explaining everyday reasoning is that they contain enough information to express, in a very simple way, what we require with respect to default information. The principal idea is that a default statement of the type ‘F’s are normally G’s’ can be expressed by saying that ‘if something is an F then it is less expected that it is non-G than that it is G’. This formulation is immediately representable in an expectation ordering $<$ by assuming that the relation $Fx \rightarrow \neg Gx < Fx \rightarrow Gx$ holds for all individuals x .

A limitation of Gärdenfors and Makinson’s work (1992, 1994) is that it does not account for the cognitive origins of an expectation ordering. The purpose of this paper is to explain the roots of the ordering. We will focus on how the information a person has about *category structure* influences her expectations about the properties that objects might have. We will use the theory of conceptual spaces (CS) to model such information (Gärdenfors 2000, 2014). After presenting these structures in Sect. 2, we show in Sect. 3 how expectation orderings are expressible in terms of prototypes and distances in CS. Section 4 brings out relations to nonmonotonic reasoning. Section 5 shows that prototypes can also be used to update expectations orderings; and Sect. 6 that the default rules that have been used in many areas of nonmonotonic logic can be derived from the expectation orderings.

2 Concept Representation in Conceptual Spaces

In brief, our position is that a proper explication of the role of expectations in reasoning must rely on a model of the structure of background knowledge. Even though this last notion has played a central role in several areas of philosophy and logic over the last decades, few efforts have been made to properly define it. In general, these disciplines worked under the assumption that both implicit and explicit knowledge can be represented in a proposition-like format in some kind of ‘belief-box’ of the cognitive agent.

Within AI, the issue has been taken more seriously. Researchers in this area have quickly noticed that the problem of the format of representation of knowledge was crucial for modeling processes like reasoning in artificial systems (cf. Lakemeyer and Nebel 1994; Woods 1987). The propositional view, widely popular at the beginnings of AI thanks to the prevalence of logical models, lost popularity for a simple reason: classical logic has several computational disadvantages when used as a framework for knowledge representation (see Minsky 1991; Sowa 1999).² A well-known example of this is the so-called frame problem in robotics (see Hayes 1988).

New representational frameworks were developed for organizing knowledge structures more efficiently. Frames (Minsky 1975), semantic networks (Quillian 1967), and conceptual graphs (Sowa 1991) were all non-propositional alternatives to classical logic that proved themselves as useful ways for representing information. However, in most cases, neither these methods have psychologically realistic foundations nor they explain where this semantic organization comes from (see Brachman 1977). In

² See also Hayes (1977), Moore (1982), or Lifschitz et al. (2008) for a discussion on this matter.

what follows, we will introduce conceptual spaces, an already well-known method for knowledge representation with solid cognitive foundations. We will use it to generate expectation orderings with the properties we are aiming for.

2.1 Conceptual Spaces

Conceptual spaces (Gärdenfors 2000, 2014) is a research program in cognitive semantics and knowledge representation studying the structure of concepts and their interrelations with geometrical and topological methods. This approach builds on two central ideas about the composition and structure of concepts and properties: (1) they are composed by clusters of quality dimensions, many of which are generated by sensory inputs; (2) they have a geometric or topological structure that is the result of the integration of the specific geometrical structures of the dimensions.

Quality dimensions can be *integral* or *separable*. They are integral when you cannot assign to an object a value in one dimension without assigning another value in another dimension (Maddox 1992). For instance, it's not possible to attribute a value to the pitch of a tone without attributing one to its loudness. When quality dimensions are not integral, they are called separable.

We define the notion of *domain* as a set of integral dimensions that are separable from all other dimensions. For instance, color properties are composed by three fundamental parameters of color perception: *hue*, *saturation* and *brightness*. Hue can be represented as a circle (as in the traditional *color circle*), saturation corresponds to an interval of the real line, and brightness (which varies from white to black) is a linear dimension with endpoints (see Gärdenfors 2000, Sec. 1.3). We can analyze any instance of color perception as a mapping that attributes some specific values to hue, intensity and brightness. More generally, we can delimit different colors as sets of possible values of these three parameters. If we interpolate this to the geometrical interpretation, then we will have a color domain that can be divided into different regions that correspond to different color properties (see Fig. 1).

A central definition in this theory, named 'criterion P' (Gärdenfors 2000, p. 71) is that natural properties (like colors) correspond to convex regions of a single domain. A region is convex when for every pair of points x and y in the region, all points between them are also in the region. In this way, criterion P assumes that the notion of *betweenness* is meaningful for the relevant domain.

Now, most predicates in natural language cannot be defined within a single domain, but as clusters of properties. This fact led Gärdenfors to make the distinction between *properties* and *concepts*. While properties are convex regions of single domains, concepts are convex regions within set of interconnected domains. This is called 'Criterion C' (Gärdenfors 2000, Sec. 4.2.1). Furthermore, for most concepts, the domains that compose them can be correlated in different ways. For example, in the case of the concept *fruit*, properties like *size* and *weight*, or *ripeness*, *color*, and *taste* co-vary. These co-variations generate expectations that are crucial for inferential procedures that exploit semantic properties.

We can define now the central notion of *conceptual space* as a collection of one or more domains with a distance function (metric) that represents properties, concepts,

Fig. 1 Color space; the property *red* is represented as a convex region that corresponds to specific values of hue, saturation and brightness

Fig. 2 ‘Fruit space’. The dotted lines represent correlations between properties for the concept *apple*

and their similarity relations. The distance function can vary; the most common one is the Euclidean, but also Manhattan, Minkowski and polar metrics may be appropriate in different contexts (see Shepard 1964; Johannesson 2002; Gärdenfors 2014).

Within this framework, objects are seen as instances of concepts and are mapped into points of the space, and concepts are represented as regions (sets of points). Similarity among concepts and objects can then be easily estimated since it is a monotonically decreasing function of their distance within the space (Shepard 1987). As an example, consider a basic conceptual space for *fruit* defined as a composition of properties of fruits from the domains of color, taste, ripeness, texture, and shape. The ‘fruit space’ will be a subset of the Cartesian product of these five domains. And the concept *apple* will occupy specific subregions of these domains that correspond to the possible properties of these fruits, as it is represented in Fig. 2.

An important advantage of representing concepts in this way is that it allows us to account for the prototypical structure of categories in a natural way (Rosch 1975, 1983; Lakoff 2008; Gärdenfors 2000). Defined as convex regions within n -dimensional spaces, a certain point in each region can be interpreted as the prototype for the property or concept. In the opposite direction, given a set of prototypes $p_1, p_2, \dots, p_n$, and a

Fig. 3 Voronoi partitioning in a 2-dimensional space

Euclidean metric, a set of n concepts can be delimited by partitioning the space into convex regions such that for each point $x \in C_i$, $d(x, p_i) < d(x, p_j)$ when $i \neq j$. This partitioning is the so-called Voronoi tessellation, a two-dimensional example of which is illustrated in Fig. 3. Thus, assuming that a metric is defined on the subspace that is subject to categorization, a set of prototypes will by this method generate a unique partitioning of the subspace into convex regions.³

This allows to represent graded membership and degrees of typicality (Rosch et al. 1976; Hampton 2007), that is, we can represent objects in the space as being more or less typical representants of the categories according to their position relative to the prototype. Representing typicality in this way has several advantages in terms of cognitive economy for processes like categorization (Gärdenfors 2000, 2014 and inductive inference (Gärdenfors and Stephens 2018; Osta-Vélez and Gärdenfors 2020). As we will see in the next section, this fact is of crucial importance while representing expectations. It should be noted that in this framework similarity is inversely correlated to distances in a CS. This makes it different from Tversky's influential notion of similarity (Tversky 1977), which is based on comparing the number of properties that two objects have in common with the number of properties where they differ.

3 Expectations in Conceptual Spaces

As mentioned earlier, our expectations about the world mirror aspects of the organization of our background knowledge. For instance, if we know that Maria bought a new pet, we would expect it to be a typical one like a dog or a cat, and, to a lesser extent, a bird or a fish. We would be surprised if we then learn that Maria's new pet is a lizard since it is rather rare to think of these animals within the scope of domestic pets. However, this could easily change if we move to contexts in which lizards are common pets. Expectations, just as typicality, are strongly dependent on cultural contexts (Schwanenflugel and Rey 1986; Lin et al. 1990).

This example shows two important features of expectations: (1) it is a 'graded' and subjective phenomenon, and (2) the strength of an expectation depends, to a significant

³ The Voronoi tessellation works also for other kinds of metrics (see Okabe et al. 1992).

extent, on the prototypical structure of concepts. To be more precise about the second point, when we receive some information about an object falling under some category, we tend to decrease our remaining uncertainty by implicitly assuming that the object also has the typical properties that are associated with the category in question. The effects of this will depend on how we represent the category and its semantic associates. As we will show in what follows, the theory of conceptual spaces can give us some useful tool for articulating such relations.

3.1 A Gricean Principle for Generating Expectations

Our analysis of expectations focuses on the agent's representation of categories as associated with clusters of properties. As we said before, the main idea is that when we categorize an object x as C , we implicitly form expectations about properties that this object is supposed to have because of falling under C . Depending on the strength of the expectations, the properties can be ordered, thus generating an expectation ordering. The organization of this ordering will depend upon the prototypical structure of C , that can be represented in its corresponding CS.

The underlying rationale for this method of generating an expectation ordering is a version of the Gricean principle of *maximal informativeness* (Grice 1975). If you are informed that an object x should be categorized as, for example, a bird, but you do not know more about what kind of bird x is, then you expect that x has all the prototypical properties of birds: You expect that x has wings, that it has a beak, that it builds nests, that it sings, that it flies, and so on. The principle of maximal informativeness says that your informant should have communicated something more specific if these expectations about x are not fulfilled.

Furthermore, when new information is added, expectations are restructured. If after learning that x is a bird you learn that it is an ostrich, you will no longer expect that it flies, nor that it sings. Instead, some new expectations will be added, such that x is big, that it runs fast and that it kicks hard. Understanding how expectations are generated and organized when some information is received, and how they are restructured when new information is added, are the two main aspects that a model of this phenomenon must account for.

3.2 Conceptual Space-Based Expectation Orderings

Let's now give some formal structure to these ideas. The conceptual space of a concept M —written $\mathcal{C}(M)$ —can be seen as a subset of the Cartesian product of n domains:

$$\mathcal{C}(M) \subseteq D_1 \times D_2 \times \cdots \times D_n$$

A crucial assumption for this framework is that an object x falling under M is represented as a n -dimensional point $x = \langle x_1, x_2, \dots, x_n \rangle \in \mathcal{C}(M)$. Each x_i represents the coordinates of x in the domain D_i , which will typically fall under some subregion $R_j \subseteq D_i$ corresponding to a subordinate property or concept of D_i .

Our main claim is that the expectations towards a sentence “ $M(x)$ ” are structured around the prototype p^M of $\mathcal{C}(M)$, which is also a point $p^M = \langle p_1^M, p_2^M, \dots, p_n^M \rangle \in \mathcal{C}(M)$.⁴ In other words, if the only thing we know about x is that it falls under concept M , we will expect it to be (close to) p^M , i.e., to have all the properties of the prototype.⁵

Now, our expectations towards the sentence “ $M(x)$ ” go well beyond the specific properties determined by the prototype.⁶ They extend to all the possible properties that an object falling under M may have. In the CS framework, this means that representing an object under a concept M implies that it may occupy any possible position in the space $\mathcal{C}(M)$. Different positions imply different properties for the object. The properties that do not apply to p^M can be considered as secondary expectations, since they are weaker (more defeasible) than the ones that apply to p^M . In general, for any property R in $\mathcal{C}(M)$, its *degree of defeasibility* can be specified as a function of its position relative to the prototype of M .

The question is now if it is possible to construct an ordering of properties that reflects their ‘expectedness level’ (and thus, their degree of defeasibility). One way of doing this is by measuring the distance to the closest point where the property is not satisfied. We can use the distance function of the conceptual space to obtain this kind of information. It turns out that one must distinguish between prototypical properties R for the category M ,⁷ that is, properties for which it holds that $R(p^M)$ and non-prototypical properties (for which it does not hold).⁸ We will talk about the ‘typicality degree’ of a property R in $\mathcal{C}(M)$ —written “ $T_M(R)$ ”—as a measure of its expectedness level, and propose the following criteria to determine it:

Typicality Criterion (TC)

- (a) For any two prototypical properties R_i, R_k in a conceptual space $\mathcal{C}(M)$, $T_M(R_i) \geq T_M(R_k)$ iff for all $x \in \mathcal{C}(M)$ with $\neg R_i(x)$, there exists $y \in \mathcal{C}(M)$ such that $\neg R_k(y)$ and $d(y, p^M) \leq d(x, p^M)$.
- (b) For any two non-prototypical properties R_i, R_k in a conceptual space $\mathcal{C}(M)$, $T_M(R_i) \geq T_M(R_k)$ iff $T_M(\neg R_k) \geq T_M(\neg R_i)$.⁹
- (c) A prototypical property always has a higher typicality degree than a non-prototypical property.

⁴ The assumption that concepts have unique prototypes is an idealization that we use here for the sake of simplicity. A more psychologically realistic model should include the possibility of concepts and properties having multiple prototypes (see Regier et al. 2005; Douven et al. 2013).

⁵ Formally, if $p_i^M \in R_j \subseteq D_i$, then we expect that $x_i \in R_j \subseteq D_i$.

⁶ We will use the same notation for talking about concepts and their lexical counterpart (predicates) in natural language.

⁷ For any point $x \in \mathcal{C}(M)$ and property $R \subseteq D_i \subseteq \mathcal{C}(M)$, “ $R(x)$ ” means that the coordinate corresponding to dimension D_i falls under the subregion corresponding to R .

⁸ We are grateful to an anonymous reviewer for pointing out a shortcoming of a previous definition.

⁹ Another way to express this is: For any two non-prototypical properties R_i, R_k in a conceptual space $\mathcal{C}(M)$, $T_M(R_i) \geq T_M(R_k)$ iff for all $y \in \mathcal{C}(M)$ with $R_k(y)$, there exists $x \in \mathcal{C}(M)$ and $R_i(x)$ such that $d(x, p^M) \leq d(y, p^M)$.

The criterion measures the distance of a region to a prototype via its closest point.¹⁰ It is important to note that in TC, we do not count numbers of instances, but the criterion is based on similarity to the prototype. In other words, our model is not probabilistic. Probabilistic models will not give the right results for expectation orderings since some property that is probable may be very atypical.¹¹

TC gives us a way to determine the typicality degree of every property in a conceptual space. Not only does it allow us to compare non-prototypical properties, but also prototypical ones. The fact that the latter are included in the prototype may lead us to believe that they have the same degree of typicality, but this is not usually the case. For instance, *having feathers*, *building nests*, and *flying* are all prototypical properties of *bird*. However, *flying* is the less typical (and thus more defeasable) of all three, since instances of birds that do not fly are more common than instances that do not build nests or have feathers. This means that the former two properties are going to have priority over the latter in an expectation order.

Considering the above we claim that, for all the properties in a the conceptual space of M , TC will produce an expectation ordering $Exp(M) = \{R_1 \geq R_2 \geq \dots \geq R_m\}$ when the following criterion is applied: given two properties R_i, R_k in $\mathcal{C}(M)$, R_i is more expected than R_k (i.e. $R_i \geq R_k$) iff R_i is more typical than R_k . This ordering of properties can be turned into an ordering of atomic sentences by saying that, for all individuals x , $R_i(x) \geq R_k(x)$ iff $R_i \geq R_k$.¹²

To see an example of how this could work, consider the fruit space introduced in the previous section, with *color*, *taste*, *ripeness*, *texture* and *shape* as dimensions. If we are told that a is an apple, our maximal expectations will be that a has the properties of a prototypical apple like being red, sweet, ripe, smooth, and round. But these properties have different degrees of typicality even if they are all in the prototype. For instance, being round is more typical than being red for apples, since it is more surprising to find a non-round apple than a non-red one. This means that points representing non-red apples are going to be closer to the prototype than points representing non-round apples in the CS. Similarly, bitterness is an odd property for apples, certainly less expected than being yellow. Thus, instances of yellow apples are going to be closer than the prototype than instances of bitter apples (see Fig. 4). An expectations ordering of properties for Apple could thus look like this: $Exp(Apple) = \{round > red > sweet > smooth > ripe > green > \dots > yellow > \dots > bitter > \dots\}$.

The typicality criterion produces a fine-grained order of expectations that makes it possible to compare individual properties. Thus, TC solves the problem of the origin of the expectation ordering that was mentioned above. We will make use of this advantage in Sect. 6.

¹⁰ Lewis and Lawry (2016) also use this distance measure for sets. There are, however, other possibilities to define the expectation ordering between properties, for example by using average distances between a prototype and a region. It is a matter of empirical research to determine which method gives the results that best fit with how humans reason.

¹¹ An example of such a model is Lieto and Pozzato (2019). We will return to a comparison with their model in Sect. 6.2.

¹² Note that we may write $R_i > R_k$ as an abbreviation for $R_i \geq R_k$ and $R_k \not\geq R_i$.

Fig. 4 Illustration of an apple region of the fruit space with points representing more and less typical instances of apples. Point y represents the closest instance to the prototype which does not satisfy the property *round*, and x represents an instance that does not satisfy the property *red*. Since $d(x, p^{Apple}) < d(y, p^{Apple})$, then $T_{Apple}(Round) > T_{Apple}(Red)$

It should be noted that, since it is based on a distance function, the TC generates a total ordering of individual properties for a given conceptual space.¹³ In many cases, however, this assumption is cognitively unrealistic. For example, which is the most typical property of an ostrich: that it runs fast or that it kicks hard? It would therefore be more natural to abstain from judging which property is more typical. In cases like this, the assumption of a unique distance measure on the conceptual space is too strong. One could extend it allowing for vague distances in the way proposed by Douven et al. (2013). This would mean that the expectation ordering would be just a partial ordering of properties. In our opinion, which ordering to use when analyzing expectation is basically an empirical question. In Sect. 7, we will discuss some aspects of this problem.

4 Relations to Nonmonotonic Logic

An ordering of properties can thus be seen as an ordering of a subset of atomic sentences within a logical language. This will allow us to connect this treatment of expectations with the results concerning nonmonotonic logic in Gärdenfors and Makinson (1994) (see also Gärdenfors 1992).¹⁴ The key idea is that α *nonmonotonically entails* β —

¹³ Proof: For any two prototypical properties R_i, R_k , if not $T_M(R_i) \geq T_M(R_k)$, then there exists $x \in \mathcal{C}(M)$ with $\neg R_i(x)$ such that for all $y \in \mathcal{C}(M)$ for which $\neg R_k(y)$, we have $d(y, p^M) > d(x, p^M)$. It follows that for all $y \in \mathcal{C}(M)$ for which $\neg R_k(y)$, there exists an $x \in \mathcal{C}(M)$ with $\neg R_i(x)$ and $d(y, p^M) > d(x, p^M)$. From this it follows that $T_M(R_k) \geq T_M(R_i)$. For any two non-prototypical properties R_i, R_k it follows that either $T_M(R_i) \geq T_M(R_k)$ or $T_M(R_i) \leq T_M(R_k)$ since either $T_M(\neg R_k) \geq T_M(\neg R_i)$ or $T_M(\neg R_k) \leq T_M(\neg R_i)$. If R_i is prototypical and R_k is non-prototypical, then, by criterion (c) of TC, it always holds that $T_M(R_i) \geq T_M(R_k)$. This concludes the proof.

¹⁴ In this section we work with a language based on propositional logic.

written $\alpha \sim \beta$ —means that β follows from α , together with all the propositions that are ‘sufficiently well expected’ in the light of α . In the context of non-monotonic logic, expectations are expressed as propositions. Using the CS approach, these expectations would be those that are associated with the prototype, as it was described in the previous section. There may, however, exist other relevant propositions that are not determined by the properties of the prototype.

A natural idea to technically specify what is meant by ‘sufficiently well’ is to require that any added sentence α be more expected than $\neg\alpha$ in the expectation ordering. It can be proved that it follows from the postulates (E1)–(E3) below that this set is consistent and that it is a maximal set with this property (see Gärdenfors and Makinson 1994). This ensures that the set of extra premises added as expectations will not contradict α and that as many as possible of the expectations are included as extra premises. This is the motivation for the following definition.

Definition 1 $C \sim$ is an expectation inference relation iff there is an ordering $\leq$ such that the following condition holds: $\alpha \sim \gamma$ iff $\gamma \in Cn(\{\alpha\} \cup \{\beta : \neg\alpha < \beta\})$

Here, Cn denotes the set of logical consequences of the premises in the set.

Gärdenfors and Makinson (1994) assume that the expectation ordering $\leq$ (which they call an ordering of epistemic entrenchment) satisfies the following properties:

- (E1) If $\alpha \leq \beta$ and $\beta \leq \gamma$, then $\alpha \leq \gamma$ (Transitivity)
- (E2) If $\alpha \sim \beta$, then $\alpha \leq \beta$ (Dominance)
- (E3) For any α and β , $\alpha \leq \alpha \wedge \beta$ or $\beta \leq \alpha \wedge \beta$ (Conjunctiveness)

(E2) means that if β is a logical consequence of α , then β is always more expected than α . (E1)–(E3) entails that $\leq$ is a total ordering, that is, either $\alpha \leq \beta$ or $\beta \leq \alpha$.

On the basis of these postulates and the definition, Gärdenfors and Makinson (1994) proved that the nonmonotonic entailment relation $\sim$ satisfies exactly the following postulates:

- If $\alpha \vdash \beta$, then $\alpha \sim \beta$ (Supraclassicality),
- If $\vdash \alpha \leftrightarrow \beta$ and $\alpha \sim \gamma$, then $\beta \sim \gamma$ (Left Logical Equivalence),
- If $\vdash \beta \rightarrow \gamma$ and $\alpha \sim \beta$, then $\alpha \sim \gamma$ (Right Weakening),
- If $\alpha \sim \beta$ and $\alpha \sim \gamma$, then $\alpha \sim \beta \wedge \gamma$ (And),
- If $\alpha \sim \beta$, then $\vdash \alpha \rightarrow \beta$ (Weak Conditionalization),
- If $\not\vdash \neg\alpha$ and $\vdash \alpha \rightarrow \beta$, then $\alpha \sim \beta$ (Weak Rational Monotony),
- If $\alpha \sim \perp$, then $\alpha \vdash \perp$ (Consistency Preservation),
- If $\alpha \sim \beta$ and $\beta \vdash \alpha$, then $\alpha \sim \gamma$ iff $\beta \sim \gamma$ (Cumulativity),
- If $\alpha \sim \gamma$ and $\beta \sim \gamma$, then $\alpha \vee \beta \sim \gamma$ (Or),
- If $\alpha \not\vdash \neg\beta$ and $\alpha \sim \gamma$, then $\alpha \wedge \beta \sim \gamma$ (Rational Monotony).

Supraclassicality means that the nonmonotonic inference relation is an extension of the classical inference relation where only explicit premises are assumed. However, the hidden expectations, which are expressed by the underlying expectation ordering, generates an inference relation that extends the classical. If the assumption that the expectation ordering is a total ordering is given up so that the expectation ordering is just assumed to be partial, then the last postulate (Rational Monotony) is no longer

satisfied. Nevertheless, the inference relation $\vdash$ has many of the properties that are desired for rational inference.

From a logical point of view, an agent would be able to draw more conclusions based on expectations if the underlying ordering is total. From a cognitive point of view, however, such an assumption may be too strong. The agent may simply lack sufficient knowledge to be able to compare two partial vectors of information about an object. For example, how can I compare the expectations about the size of an apple to expectations about the color of its seeds? In general, if two domains in a conceptual space are largely independent, then it may be difficult to compare values in one domain to those of the other. Consequently, the resulting expectation ordering should only be partial.

5 Criteria for Updating Expectations Orderings

As we mentioned earlier, expectation orderings are context-specific. As such, they are dynamic structures that change according to the information available. A central problem for the study of expectation-based inference is to understand the principles according to which these orderings are updated when new information is added. In what follows, we will show how the CS approach may bring some light on this issue. In general, different kinds of information, for example perceptual information, may generate different kinds of updates. In this article, however, we will restrict our analysis to updates produced by information that is already present in the ordering generated by TC.

First of all, notice that only information that adds specificity to a previous informational state will change the expectation orderings. Obviously, if we are told that b is a dog and later that b is a mammal, then the expectation ordering remains the same since the new information is already implicit in the initial informational state. According to the framework defended in this article, reasoning under uncertainty about properties of objects amounts to specifying the position of a point in a n -dimensional space. Each new piece of information with a more specific value in one of the dimensions reduces the space of uncertainty.

5.1 Prototype Shifts

We start by analyzing a kind of updating we name ‘prototype shift’. The idea is that given an expectation ordering $Exp(M)$ structured around prototype p^M , if we are told that x is also G and we know that G is a subordinate concept of M ($G \subset M$), then $Exp(M)$ will be updated into $Exp(M \& G)$, changing its prototype to p^G and being re-structured accordingly (following the two criteria explained in Sect. 3). This is supposed to reduce the number of properties in the original expectation ordering since the new information will “shrink” the conceptual space.

To see an example, suppose that we are told that a is a bird. The associated expectation set $Exp(BIRD)$ will contain a large number of color properties and shapes (the bird category has a large color and shape variability). If we are then told that a is a

penguin, the updated expectation ordering $Exp(BIRD \& PENGUIN)$ will change its previous prototype to the prototype of *penguin* and it will lose all the elements referring to colors which are not black and white, as well as the elements referring to not penguin-like shapes. Also, it will lose the property of flying, in the dimension encoding information about locomotion.

5.2 Updating Via Properties

A less dramatic kind of update happens when information about specific properties is added. Given the information $M(x)$ and the associated expectation ordering $Exp(M)$, if we get to know that $F(a)$, and F is a property (subregion) of the domain D_i in $\mathcal{C}(M)$, then the ordering $Exp(M \& F)$ will be equal to $Exp(M)$ minus all the properties that are incompatible with F in D_i . If D_i is partitioned into disjoint regions, then $Exp(M \& F)$ will not include all the other properties of D_i which are not F . If this is not the case, then $Exp(M \& F)$ might lose or might not lose properties from $Exp(M)$, depending on the structure of D_i .

Consider the simple example in which we start with the information $Man(a)$ and we get to know later also $Bachelor(a)$. Then $Exp(MAN)$ will shrink by losing the property *married* (previously in $Exp(MAN)$). However, if we later get to know that a plays golf as a hobby, then we cannot delete from $Exp(MAN \& BACHELOR)$ all the other properties corresponding to the hobby dimension: since this domain is not disjoint, it might be the case that a man has several hobbies. Thus, for properties corresponding to not-disjoint dimensions, the update will depend on the particular structure of the domain in question.

5.3 Updating Via Properties and Correlations

A more complex case of *updating via properties* happens when the new property is correlated to another property in another domain. For instance and as said before, in the fruit space, dimensions like ripeness and texture, or color and taste are strongly correlated, such that apples are commonly expected to be sweet, but green apples are expected to be sour, since these two properties are correlated.

For this kind of cases, the updating procedure will be the following. Given $Exp(M)$ and two correlated properties G and F from domains D_i and D_k in $\mathcal{C}(M)$. When new information G is added, we will have that:

- (i) $Exp(M)$ will be updated via property G in the previously explained way;
- (ii) $Exp(M \& G)$ will have F as maximal or close to the maximal;
- (iii) For any other $H \subseteq D_k$, F will be more expected than H in $Exp(M \& G)$.

To continue with the apple example, suppose that our initial set is $Exp(APPLE) = \{round > red > sweet > smooth > green > ripe > \dots > yellow > \dots > bitter > \dots\}$. If we are then told that apple a is green, properties like *red* and *yellow* will be deleted, while *sour* will approximate the maximal and be more expected than any other taste-related property. The updated set would look like this:

$Exp(APPLE \& GREEN) = \{green > round > sour > smooth > ripe > sweet > \dots > bitter > \dots\}.$

6 Default Rules, Degrees of Defeasibility and Inferential Strength

6.1 Generating Default Rules

Everyday reasoning rarely follows a set of formal deductive rules, as suggested by the traditional logicist views. Instead, it exploits semantic information from our conceptual system in the form of concept-specific rules of inference (cf. Osta-Vélez and Gärdenfors 2020; Sperber and Wilson 1986). A formalization of this idea was developed by default logics, through the notion of *default rules* (Brewka et al. 2007; Delgrande 2011; Horty 2012). Default rules are predicate-specific inference rules. They are meant to capture facts that are normally or typically the case when something is claimed to fall under some predicate M . For instance, the fact that most mammals have fur can be captured by the default rule $Mammal(x) \rightarrow HaveFur(x)$.¹⁵ Defaults like this can be used as rules of inference in nonmonotonic logical systems to extend derivations.

A foundational problem of default logics concerns the interpretation of the notion of a *default rule* (Delgrande 2011). Two epistemological issues affect this notion: (1) it is not clear where defaults come from; and (2) before multiple defaults that can be applied to the same object, a decision method has to be applied to determine which defaults have priority over the others (see Horty 2012, p.19). A well-known example is the ‘Nixon diamond’: By default, quakers are pacifists and, by default, republicans are not pacifist; however, Richard Nixon is both a quaker and a republican. Thus, when determining to conclude whether Nixon is a pacifist or not, one must violate one of the default rules. Gärdenfors and Makinson (1994, Section 3.3) argued that expectation orderings offer a natural way of representing defaults. In what follows, we will adapt and develop this idea for CS-based orderings.

A natural answer to (1) is that defaults express strong regularities about phenomena that we can use as ‘inference tickets’ in reasoning. However, notice that if we focus on cognitive modeling, we should prioritize an internalist interpretation: the idea that defaults express close conceptual relationships codified in our background knowledge. In this sense, the question about the origins of defaults becomes a question about how they emerge from the structure of this background knowledge. As suggested by Gärdenfors (2000, p. 117), the CS framework offers an answer to this question. Again, when we categorize an object under a concept, we automatically represent it as having a cluster of properties associated with the concept’s prototype. For instance, categorizing something as a fruit, implies also categorizing it as sweet. This feature about categorization can also explain the kind of conceptual transitions that are expressed by default rules. An answer to the problem about the origins of defaults can be found, then, in the explanation that CS offer about the prototypical structure of concepts and its exploitation in inference and categorization.

¹⁵ We use expressions like $F(x) \rightarrow G(x)$ as an horizontal abbreviation of *normal* default rules, generally written as $(F(x) : G(x)/G(x))$. These are not formulas, but schemes for all their ground instances.

Let's turn now to (2). Notice that a default rule of the form '*F*'s are normally *G*'s' can be expressed in terms of expectations as 'if something is an *F* then it is less expected that it is non-*G* than that it is a *G*'. This can be represented in an expectation ordering as $F(b) \rightarrow \neg G(b) < F(b) \rightarrow G(b)$ for all *b*. Since the consequence relation $\alpha \sim \beta$ is equivalent to $\alpha \vdash \beta$ or $\alpha \rightarrow \neg\beta < \alpha \rightarrow \beta$ (Gärdenfors and Makinson 1994, p. 214), the fact that $F(b) \sim G(b)$ can be directly proved. This means that an expectation-based inference relation is a nonmonotonic relation with *embedded* defaults. Thus, we don't need any explicit representation of them.

Furthermore, TC gives us information about the comparative strength of each of these defaults. For instance, for the concept *bird*, it is to be expected that (according to TC) the property *build nests* has priority over *fly*. This means that the default rule $Bird(x) \rightarrow Fly(x)$ will be weaker than $Bird(x) \rightarrow BuildNest(x)$. In general, the more entrenched a property is in the CS-generated ordering, the stronger its correspondent default rule will be.

The discussion above shows that the framework introduced in this article provides new insights for the two fundamental epistemological problems of default logic and suggests constructive solutions to them.

6.2 Typicality and Conjunction Fallacy

As we just explained, one of the central aims of default logics is to capture our use of typical predicate-relations in nonmonotonic reasoning. In recent years, various formal models have been developed in order to tackle in a more systematic way the role of typicality in reasoning. For instance, Propositional Typicality Logic (Booth et al. 2013) uses a classical propositional language plus a typicality operator for specifying the set of typical situations in which some formula holds. Description Logic has been also used for modeling this phenomenon (see Giordano et al. 2008). In particular, Lieto and Pozzato (2018, 2019) developed a rich formal framework that uses Description Logic to account for typicality in nonmonotonic reasoning and conceptual combination.

The model presented here is a natural framework for accounting for the role of typicality in reasoning. We will not offer a systematic explanation of this now, but we will give an idea of how this can be done by offering an expectation-based explanation of the conjunction fallacy.

As is well known, Tversky and Kahneman (1974, 1983) showed that intuitive judgments of probability do not mirror the principles of standard probability calculus. They argued that in many cases, people violate these principles because they prioritize intuitive heuristics that exploit typicality relations for estimating the probability that an object has a certain property. Their claims are supported by a famous experimental case called the 'Linda problem'. In brief, subjects were given the following information: Linda is 31 years old, single, outspoken, and very bright. She has a degree in philosophy, and while studying, she participated in anti-nuclear demonstrations and was involved with issues of discrimination and social justice. Afterwards, subjects were asked which of the following statement is more probable:

- (i) Linda is a bank teller.
- (ii) Linda is a feminist bank teller.

A large majority of the subjects answered that sentence (ii) is more probable, although by the laws of probability (i) is at least as probable as (ii).

The issue is that ordinary people do not interpret the word ‘probable’ in the same way as people educated in statistics. The paradox disappears when formulated in frequentist terms (see Cosmides and Tooby 1996). Our Gricean analysis coheres with Tversky and Kahneman idea that elements of typicality come into play when making the judgment.

According to our approach, when we are given the initial description of Linda, we construct a partial vector of information that approximately locate her within a sub-region of a multidimensional person-space. Following the terminology used in earlier sections, we represent Linda as an object $l = \langle x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, \dots \rangle \in \mathcal{C}(\text{Person})$, with the first 8 coordinates falling under subregions of the properties in the description:¹⁶

$x_1 \in \text{Woman}$	$x_5 \in \text{Enviromentalist}$
$x_2 \in \text{Young}$	$x_6 \in \text{Socially Aware}$
$x_3 \in \text{Single}$	$x_7 \in \text{Bright}$
$x_4 \in \text{Studied Philosophy}$	$x_8 \in \text{Outspoken}$

The ellipsis in the set stands for the dimensions of the person-space we do not have information about (we cannot assign a property or ‘value’ to them). Conforming to our analysis, if required, these are going to be filled out by properties from the expectation set evoked by l .

Now the question is: Why do we tend to believe that it is more likely for l to be a feminist bank teller than a bank teller? Our answer is that since some of the properties of l are highly correlated to the property *feminist* and negatively correlated to the property *bank teller*, l will be located closer to the composite prototype of *feminist bank teller* (“ p^{FBT} ”) than to the prototype of *bank teller* (“ p^{BT} ”) in the person-space. In other words, since $d(p^{FBT}, l) < d(p^{BT}, l)$, feminist bank teller will be more typical (and thus more expected) for l than bank teller.

The first one to analyze the conjunction fallacy in terms of defaults or expectations seems to be Veltman (1998).¹⁷ He defines a collection X of properties to be more representative of the category A than the category B if (i), for some properties P in X , P is a ‘normal’ property of A but only an exceptional property of B , and (ii) no property P in X is a ‘normal’ property of B but only an exceptional property of A . Given this definition, the properties X ascribed to Linda are more representative of the category *feminist bank tellers* than of the category *bank tellers*. Veltman’s solution is similar to ours in that he also bases it on the typicality of properties for a category. However, he does neither exploit conceptual spaces, nor any form of distance measure.

¹⁶ For the sake of argument, we assume that there are (more or less) precise categories in non-overlapping dimensions for the properties described in the person conceptual space.

¹⁷ We are grateful to Rineke Verbrugge for bringing this article to our attention.

Lewis and Lawry (2016) and Lieto and Pozzato (2019) also analyze the conjunction fallacy with the aid of their respective models of concept combination so it is interesting to compare with their solutions. The goal of the model presented by Lewis and Lawry (2016) is to represent hierarchies of concepts. Their model is similar to ours in that they also use a geometrical approach based on conceptual spaces. Instead of using Voronoi tessellations to determine category membership they use Random set theory. Nevertheless, since their explanation of the conjunction fallacy is also based on distances to prototypes, it is similar in spirit to ours.

The description logic presented by Lieto and Pozzato (2019) is less similar since their representations are probabilistic rather than geometric. Therefore, it is difficult to compare their account of the conjunction fallacy to the solution we present. There are, nevertheless interesting similarities between the models: The logic for their typicality operator **T** can be shown to be equivalent to the nonmonotonic logic presented in Sect. 4 of this paper.¹⁸

6.3 Inferential Strength

When an inference is based on defeasible premises, a criterion for judging the strength of the inference is helpful. Deductive inferences, for instance, are often considered as maximally strong since they preserve truth. However, in everyday reasoning they can be analyzed as having different strengths according to the ‘quality’ of the information in the premises. In other words, the structural properties of a deductive inference do not prevent it from being weak from the perspective of everyday reasoning.

The strength of an inductive inference, on the other hand, might depend on the size of the sample (or on the probability) of the premises. Other forms of inductive inference, like category-based induction, depend on a more complex combination of factors related to the exploitation of different conceptual relations (see Osta-Vélez and Gärdenfors 2020).

Remember that an expectation ordering is a kind of epistemic entrenchment ordering. The notion of epistemic entrenchment comes from belief revision, and it was meant to capture the idea that within a belief system, certain beliefs are more susceptible to be revised or deleted than others. In an expectation ordering, the position of a belief in the ordering gives information about its comparative degree of defeasibility (Gärdenfors and Makinson 1994, p. 209). In this article we have suggested TC as a criterion for measuring such a degree of defeasibility for properties: The more typical a property is, the less defeasible it is.

To give an illustrative example, suppose I’m told that Maria bought a new pet. I could infer that it flies if I use as implicit premise the belief from my set of expectations that the pet in question is a bird. However, this inference would be quite weak, since *bird* would typically be less entrenched in $Exp(PET)$ than concepts like *dog* or *cat*. If, instead of *bird*, I use as implicit premise that Maria’s new pet is a dog or a cat,

¹⁸ Lieto and Pozzato (2019) remark that their logic is a reformation of Lehmann and Magidor’s (1992) rational logic. Gärdenfors and Makinson (1994) showed that this logic is equivalent to the logic present in Sect. 4 above.

then I would nonmonotonically infer that it does not fly, and this inference would be clearly stronger than the other one.

The above example shows that when we reason nonmonotonically, the strength of our inferences depends on the choice of the information from the set of expectations that we use as implicit premises. In more general terms, we can say that the strength of an inference is a negative function of the degree of defeasibility of the propositions that are used as implicit premises. As we mentioned before, our framework allows for an explication of this criterion by stating that the information to be used is always the one present in the maximal point of the set of expectations, i.e. the prototypical information of the concept in question or the one that results from an updating process as defined in Sect. 4.

As it was explained in Sect. 3, even among the sufficiently well expected information in the set of expectations, some propositions are less defeasible than others, and as a result, they will produce stronger inferences. For a system of default logic, a comparative notion of inferential strength can be defined as a positive function of the strength of the default rules used in specific inferences. As we showed earlier, this strength measure comes from the position of the rule in the CS-based expectation ordering, which depends ultimately on the distance measure defined in TC. To give a simple example, an inference that uses the default $Bird(x) \rightarrow HaveFeathers(x)$ will be stronger than one that uses the default $Bird(x) \rightarrow Fly(x)$ since following TC, $HaveFeathers(x)$ will be more entrenched in the expectation order than $Fly(x)$ even if both pertain to the prototype of *bird*.

Finally, another important issue in default logics concerns the possibility of having defaults expressing conflicting information (cf. Reiter and Criscuolo 1981; Touretzky 1984). The most famous example of this situation is the ‘penguin principle’ (Lascarides and Asher 1993): if we are told that x is a bird and that it is also a penguin, then we will have at our disposal the defaults $Bird(x) \rightarrow Fly(x)$ and $Penguin(x) \rightarrow \neg Fly(x)$. If we do not have a clear criterion to choose which default has priority, we will end up addressing a contradiction. Above we mentioned the ‘Nixon diamond’ where two conflicting defaults seemingly lead to the conclusion that Nixon is both a pacifist and not a pacifist. The TC criterion gives a way of resolving the conflict, since once it is decided whether Nixon is more atypical as a quaker than as a republican, the resulting expectation ordering determines which of the default rules should yield.

7 Conclusions

Despite the significant progress made in the field of nonmonotonic reasoning during the past decades, the available logical models still suffer from various epistemological issues. In particular, several formal systems make strong assumptions about the role of background knowledge and entrenchment relations of beliefs without providing any cognitive or epistemological argument for them.

In this article we argued that some of these problems are due to the intrinsic limitations of propositional-based models for capturing the internal structure of conceptual knowledge. We showed that combining the CS framework for concepts with an expectation-based analysis of nonmonotonic inference is a fruitful way of extending

the modelling tools of these logical systems while enriching their theoretical foundations.

Above we have discussed the model suggested by Lieto and Pozzato (2018) and there are other attempts to computationally implement reasoning with default rules (e.g. Delgrande and Schaub 2000; Brewka et al. 2007). However, in these systems the default rules must be provided by the user and they do not generalize well. We submit that a CS approach to modelling reasoning with default and nonmonotonic reasoning in general is a better method when attempting to build artificial systems with these capacities. Once the domains with their concept regions and distance functions have been implemented (we are not saying that this is an easy task), then the TC principle provides a direct way to generate expectation orderings and default rules, and thereby a method to calculate nonmonotonic inferences. Implementations, however, still remain to be constructed.

The ideas defended in this article are directly related to a kind of reasoning known in psychology as category-based induction. This mechanism consists in judging the strength of arguments that project some property from one or more categories to another category, exploiting conceptual similarities (see Feeney 2017). In Osta-Vélez and Gärdenfors (2020) we introduced a model of this phenomenon based on distances on conceptual spaces, that also uses the notion of expectations. We believe that explicating both nonmonotonic reasoning and category-based induction with the same modeling framework could be of great value for the various disciplines studying these phenomena; and contribute to the systematization of the formal frameworks used in theories of reasoning.

Finally, one limitation of the research presented here is that the expectations we are able to model concern only object predication. However, expectations are pervasive in cognition and a full model of their role in reasoning should also account for more complex lexical items, like verbs or relational and functional predicates. In Gärdenfors (2014; To appear), the conceptual space framework has been extended for modeling expectations based on the structure of events, in particular causal inferences. Future work on expectation-based nonmonotonic reasoning will hopefully result in a more general model.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Booth, R., Meyer, T., & Varzinczak, I. (2013). A propositional typicality logic for extending rational consequence. *Trends in Belief Revision and Argumentation Dynamics*, 48, 123–154.

- Brachman, R. J. (1977). What's in a concept: structural foundations for semantic networks. *International Journal of Man-machine Studies*, 9(2), 127–152.
- Brewka, G., Roelofsens, F., & Serafini, L. (2007). Contextual default reasoning. In *Proceeding of IJCAI-2007* (pp. 268–273).
- Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58(1), 1–73.
- Delgrande, J. P. (2011). What's in a default? Thoughts on the nature and role of defaults in nonmonotonic reasoning. In Brewka, G., Marek, V. W., & Truszczyński, M., (eds.), *Nonmonotonic reasoning. Essays celebrating its 30th anniversary* (pp. 89–109). College Publications.
- Delgrande, J. P., & Schaub, T. (2000). Expressing preferences in default logic. *Artificial Intelligence*, 123(1–2), 41–87.
- Douven, I., Decock, L., Dietz, R., & Égré, P. (2013). Vagueness: A conceptual spaces approach. *Journal of Philosophical Logic*, 42(1), 137–160.
- Feeney, A. (2017). Forty years of progress on category-based inductive reasoning. In L. J. Ball & V. Thompson (Eds.), *International handbook of thinking and reasoning* (pp. 167–185). Routledge.
- Gärdenfors, P. (1992). The role of expectations in reasoning. In M. Masuch & L. Pólos (Eds.), *Knowledge representation and reasoning under uncertainty* (pp. 1–16). Springer-Verlag.
- Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. MIT press.
- Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. MIT Press.
- Gärdenfors, P., & Makinson, D. (1994). Nonmonotonic inference based on expectations. *Artificial Intelligence*, 65(2), 197–245.
- Gärdenfors, P., & Stephens, A. (2018). Induction and knowledge-what. *European Journal for Philosophy of Science*, 8(3), 471–491.
- Giordano, L., Gliozzi, V., Olivetti, N., & Pozzato, G. L. (2008). Reasoning about typicality in preferential description logics. In *European workshop on logics in artificial intelligence* (pp. 192–205). Springer.
- Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics* (Vol. 3, pp. 41–58). Academic Press.
- Hampton, J. (2007). Typicality, graded membership, and vagueness. *Cognitive Science*, 31(3), 355–384.
- Hayes, P. J. (1977). In defense of logic. In *Proceeding of IJCAI-77* (Vol. 1, pp. 559–565).
- Hayes, P. J. (1988). What the frame problem is and isn't. In Z. Pylyshyn (Ed.), *The Robot's dilemma: The frame problem in artificial* (pp. 123–137). Ablex Publishing Corporation.
- Horty, J. F. (2012). *Reasons as defaults*. Oxford University Press.
- Johannesson, M. (2002). *Geometric models of similarity* (Vol. 90). Lund University Cognitive Studies.
- Lakemeyer, G., & Nebel, B. (1994). Foundations of knowledge representation and reasoning. In G. Lakemeyer & B. Nebel (Eds.), *Foundations of Knowledge Representation and Reasoning* (pp. 1–12). Springer.
- Lakoff, G. (2008). *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago press.
- Lascarides, A., & Asher, N. (1993). Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy*, 16(5), 437–493.
- Lehmann, D., & Magidor, M. (1992). What does a conditional knowledge base entail? *Artificial Intelligence*, 55(1), 1–60.
- Lewis, M., & Lawry, J. (2016). Hierarchical conceptual spaces for concept combination. *Artificial Intelligence*, 237, 204–227.
- Lieto, A., & Pozzato, G. L. (2018). A description logic of typicality for conceptual combination. In M. Ceci, N. Japkowicz, J. Liu, G. Papadopoulos, & Z. Raś (Eds.), *Foundations of Intelligent Systems: International Symposium on Methodologies for Intelligent Systems* (pp. 189–199). Springer.
- Lieto, A., & Pozzato, G. L. (2019). A description logic framework for commonsense conceptual combination integrating typicality, probabilities and cognitive heuristics. *Journal of Experimental & Theoretical Artificial Intelligence*, 1–36.
- Lifschitz, V., Morgenstern, L., & Plaisted, D. (2008). *Knowledge representation and classical logic* (Vol. 3, Chap. 1, pp. 3–88). Elsevier.
- Lin, P.-J., Schwanenflugel, P., & Wisenbaker, J. (1990). Category typicality, cultural familiarity, and the development of category knowledge. *Developmental Psychology*, 26(5), 805.
- Maddox, T. (1992). Perceptual and decisional separability. In G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147–180). Lawrence Erlbaum Associates Inc.
- Minsky, M. (1975). *A framework for representing knowledge* (pp. 211–277). McGraw-Hill.

- Minsky, M. (1991). Logical versus analogical or symbolic versus connectionist or neat versus scruffy. *AI Magazine*, 12(2), 34.
- Moore, R. C. (1982). *The role of logic in knowledge representation and commonsense reasoning*. Artificial Intelligence Center: SRI International.
- Oaksford, M., & Chater, N. (2009). The uncertain reasoner: Bayes, logic, and rationality. *Behavioral and Brain Sciences*, 32(1), 105–120.
- Okabe, A., Boots, B., & Sugihara, K. (1992). *Spatial tessellations: concepts and applications of Voronoi diagrams*, (Vol. 501). John Wiley & Sons.
- Osta-Vélez, M., & Gärdenfors, P. (2020). Category-based induction in conceptual spaces. *Journal of Mathematical Psychology*, 96, 102357.
- Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral science*, 12(5), 410–430.
- Regier, T., Kay, P., & Cook, R. S. (2005). Focal colors are universal after all. In *PNAS-2005* (Vol. 102, pp. 8386–8391).
- Reiter, R., & Criscuolo, G. (1981). On interacting defaults. In *Proceeding of IJCAI-81* (Vol. 81, pp. 270–276).
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3), 192.
- Rosch, E. (1983). Prototype classification and logical classification: The two systems. In E. Scholnick (Ed.), *New trends in conceptual representation: Challenges to Piaget's theory* (pp. 73–86). Lawrence Erlbaum Associates.
- Rosch, E., Simpson, C., & Miller, S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human perception and performance*, 2(3), 491–502.
- Schwanenflugel, P. J., & Rey, M. (1986). The relationship between category typicality and concept familiarity: Evidence from spanish-and english-speaking monolinguals. *Memory & Cognition*, 14(2), 150–163.
- Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, 1(1), 54–87.
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237(4820), 1317–1323.
- Sowa, J. F. (1991). Toward the expressive power of natural language. In J. F. Sowa (Ed.), *Principles of semantic networks* (pp. 157–189). Elsevier.
- Sowa, J. F. (1999). *Knowledge representation: Logical, philosophical and computational foundations*. Brooks/Cole Publishing Co.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Harvard University Press Cambridge.
- Touretzky, D. S. (1984). Implicit ordering of defaults in inheritance systems. In *Proceedings of AAAI-84* (pp. 322–325).
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4), 293.
- Veltman, F. (1998). Een zogenaamde denkfout. *Algemeen Nederlands Tijdschrift voor Wijsbegeerte*, 90(1), 11–25.
- Woods, W. A. (1987). Knowledge representation: What's important about it? In N. Cercone & G. McCalla (Eds.), *The knowledge frontier* (pp. 44–79). Springer.