Minds and Machines
https://doi.org/10.1007/s11023-023-09640-2

Reasoning with Concepts: A Unifying Framework

Peter Gärdenfors[1] · **Matías Osta-Vélez**[2,3]

Received: 27 December 2022 / Accepted: 20 June 2023

Abstract
Over the past few decades, cognitive science has identified several forms of reasoning that make essential use of conceptual knowledge. Despite significant theoretical and empirical progress, there is still no unified framework for understanding how concepts are used in reasoning. This paper argues that the theory of conceptual spaces is capable of filling this gap. Our strategy is to demonstrate how various inference mechanisms which clearly rely on conceptual information—including similarity, typicality, and diagnosticity-based reasoning—can be modeled using principles derived from conceptual spaces. Our first topic analyzes the role of expectations in inductive reasoning and their relation to the structure of our concepts. We examine the relationship between using generic expressions in natural language and common-sense reasoning as a second topic. We propose that the strength of a generic can be described by distances between properties and prototypes in conceptual spaces. Our third topic is category-based induction. We demonstrate that the theory of conceptual spaces can serve as a comprehensive model for this type of reasoning. The final topic is analogy. We review some proposals in this area, present a taxonomy of analogical relations, and show how to model them in terms of distances in conceptual spaces. We also briefly discuss the implications of the model for reasoning with concepts in artificial systems.

Keywords Analogy · Category-based induction · Conceptual spaces · Expectations · Generics · Reasoning · Similarity · Typicality

✉ Matías Osta-Vélez
matiasosta@gmail.com

Peter Gärdenfors
Peter.Gardenfors@lucs.lu.se

[1] Department of Philosophy and Cognitive Science, Lund University, Lund, Sweden

[2] Instituto de Filosofía, FHCE, Universidad de La República, Montevideo, Uruguay

[3] Department of Philosophy, Heinrich-Heine University of Düsseldorf, Düsseldorf, Germany

Published online: 11 July 2023

1 Introduction

Concepts are often understood as the 'building blocks of thought' (e.g., Fodor, 1983; Pinker, 2007). Consequently, they are assumed to play a central role in the explanation of rational thinking. However, theories of concepts in psychology and philosophy rarely integrate theories of reasoning and vice versa. The upshot is that no scientific or philosophical story provides a systematic explanation of how concepts are involved in the processes that characterize reasoning.

The divorce between these two areas of research traces back to classical logic, which was founded upon the idea that a theory of argument validity can dispense with a theory of conceptual content. In other words, logic is presumed to be 'topic neutral' since it regards the syntax of a few logical operators as sufficient for constructing a theory of rational inference (see, MacFarlane, 2000). In psychology, the influence of this heritage is evident in what is often referred to as the 'deductivist paradigm' (or 'classical view of reasoning'). This perspective posits that inference operates through the application of abstract rules to propositionally structured information, akin to a proof-theory system (e.g., Rips, 1994; Smith et al., 1992). However, many forms of human reasoning are evidently concept-based and rely on semantic mechanisms that go beyond the scope of classical logic. In addition, experimental psychology has demonstrated that the thematic content of the problems we reason about affects both the strategies used and processing fluency; something that has proven challenging to reconcile with the syntactic paradigm (see, Pollard & Evans, 1987; Kellen & Klauer, 2020).

Despite significant theoretical and empirical progress, a unified model explaining how concepts are used in reasoning remains elusive. This article utilizes the theory of conceptual spaces (Gärdenfors, 2000, 2014) as a comprehensive framework for analyzing various concept-based inferences explored in cognitive science. The fundamental premise of our approach is that reasoning exploits the structural properties of conceptual representation, and not the syntactic properties of language—as is assumed in the classical view. Conceptual spaces enable us to model several key notions to the analysis of reasoning within psychological literature, such as prototypes, similarity, typicality, and diagnosticity. The applicability of these definitions across a range of psychological areas underscores the unifying strength of our approach.

The critical motivation behind our proposal aligns with the so-called 'new paradigm in the psychology of reasoning' (see Tessler & Goodman, 2019; Oaksford & Chater, 2020), which uses Bayesian models (and probabilities in general) to examine various aspects of reasoning and rationality. However, our approach diverges in several key aspects. Firstly, the new paradigm predominantly focuses on the use of conditionals in domain-general forms of reasoning. In contrast, the conceptual space approach centers on domain-specific inference types that depend on the extralogical terms of the language rather than the structure of the sentences. Secondly, the new paradigm does not delve into the nature of concepts or their role in reasoning, while the framework presented here bridges these two areas of inquiry. Lastly, we posit that not all mechanisms employed in reasoning

can be modeled probabilistically. For instance, we will demonstrate in the section on generics that probabilistic models fail in handling certain cases. Similarly, probabilistic models do not effectively accommodate analogical reasoning. Furthermore, people frequently violate fundamental principles of probabilites while reasoning. A classic example is the 'Linda problem' (Tversky & Kahneman, 1982), which exhibits the so-called conjunction fallacy. Osta-Vélez, M., & Gärdenfors, P. (2022a) illustrate how this example can alternatively be analyzed in terms of expectations based on similarities rather than probabilities.[1]

Our analysis will begin by examining inferential processes that make use of the prototypical structure of concepts. Thanks to the pioneering work of Rosch (1975) and Barsalou (1985), it is now widely accepted that most natural categories possess a prototypical structure. This feature of semantic representation is particularly important for cognitive tasks such as categorization, but it also plays a role in inferences under uncertainty. Our first topic will be a model of reasoning based on *expectations* about concepts. It is important to note that our notion of expectation is non-probabilistic.Our second topic concerns the role of generic expressions in concept formation and reasoning. Generics are statements of the form Xs are Ys" (e.g., "Tigers are striped") or Xs cause Y" (e.g., "Sharks kill people"), and they have been notoriously difficult to analyze using traditional logical tools. The numerous proposals in the literature all have encountered problems. We propose that generics should not be approached from a truth-conditional perspective, but should rather be regarded as sentences that express expectations that may vary in strength. This allows them to convey information about the structure of categories and to be used in conjunction with factual knowledge to draw practical conclusions. To quantify the strength of a generic expression, we suggest employing the concept of typicality which we deefine in terms of distances between properties and prototypes in conceptual spaces. Our third topic of investigation focuses on inductive reasoning. Philosophers have extensively studied induction, but it continues to present challenges, particularly for logic-based approaches. One prominent form of concept-based reasoning is category-based induction, an inferential mechanism that utilizes knowledge of conceptual relationships to estimate the likelihood of projecting a property from one category to another. For example, inferring that wolves have a certain property because dogs share the same property seems plausible due to the strong similarity between these categories. In recent decades, psychologists have identified several features of this mechanism and proposed various formal models to account for it. In our analysis, we review some of the key proposals and argue that the theory of conceptual spaces can serve as a unifying framework for modeling category-based induction.

Our final topic centers around analogy, a fundamental cognitive mechanism that helps organize our conceptual knowledge by identifying similarities between

[1] It's also worth noting that in the tradition of symbolic AI, attempts have been made to construct models that elucidate the relationship between concepts and reasoning. A notable example is Semantic Networks, which have found diverse applications in reasoning studies (see Shastri, 1989). Additionally, Nonmonotonic logic was developed for similar purposes (see Brewka, 1991). More recently, various forms of Description Logics, particularly Typicality Logic, have been designed to model how we reason with prototypes (for example, Lieto and Pozzato, 2019). However, none of these models is built on or inspired by cognitive hypotheses about conceptual structure.

seemingly disparate areas of knowledge. Analogy is widely recognized to play a critical role in reasoning (e.g., Bartha, 2010; Hofstadter & Sanders, 2013). However, the literature on analogies is quite diverse, with various theories being put forth. In our analysis, we review some of the most significant proposals in this area and present a comprehensive model that includes a taxonomy of analogical relations. Our proposed model offers a new perspective on how to conceptualize and model analogies, using the concept of distances in conceptual spaces. This approach provides a unified framework for understanding analogical reasoning and its underlying cognitive processes.

Each of the four concept-based cognitive mechanisms that we examine in this article has traditionally been studied in isolation using different approaches. However, a central theme of this article is to demonstrate that there are strong relationships between them, all of which rely on similarity, typicality and diagnosticity. We argue that by using conceptual spaces as a modeling tool, we can better understand these interrelationships. Our proposed model offers a unification of these different mechanisms, enabling us to generate new predictions about reasoning with concepts.

2 The Role of Expectations in Reasoning

Studying reasoning from the perspective of classical logic ties us to two unwarranted assumptions: 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 (see Gärdenfors, 1992). This leads us to conceive reasoning from a purely syntactic perspective without any relation to semantic notions (e.g., Bonatti, 1994). The contents of the predicates in the sentences are considered completely irrelevant (e.g., Fodor & Pylyshyn, 2015). In brief, classical logic presumes a sharp distinction between form and content and depicts reasoning as informationally conservative in that the information in the conclusion of an argument is contained in the premises.

Everyday reasoning, however, clearly builds on more than the logical form of explicit premises. Because most of our decisions are made under uncertainty, our inferential mechanisms can hardly afford to be informationally conservative (see Oaksford & Chater, 2009). We must constantly take risks and use our background knowledge in productive ways to complement the explicit information on the premises. One of our main theses is that much of this background information consists of knowledge about the structure of concepts and their relations to each other.

One particular way in which this use of background knowledge expresses itself is through our *expectations* about the world. For instance, if we see that an apple is red, we expect it to be sweet; or if we turn the ignition key of a car, we expect the engine to start. In general, our expectations about the world are crucial for guiding our reasoning and actions in everyday life, and they build directly on the structure of our background knowledge.

In the logical tradition, such expectations have been put to work in so-called nonmonotonic logic. Gärdenfors (1992) and Gärdenfors and Makinson (1994) have

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 in inferences 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. Expectations are thus 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.

This approach is limited in that it does neither say anything about how the expectations arise nor about how their strength in an argument can be gauged. We shall argue that knowledge about the structure of concepts and what is *typical* of things falling under a concept can be used to fill these gaps. For this purpose, we next introduce the theory of conceptual spaces as a tool for modelling such knowledge.

3 Conceptual Spaces as a Modelling Framework

3.1 The Basic of Conceptual Spaces

Conceptual spaces (Gärdenfors, 2000, 2014) have been developed as a research program in semantics studying the structure of concepts and their interrelations using geometrical methods. The approach builds on two central ideas about the composition and structure of concepts and properties: (i) they are composed by clusters of quality dimensions, many of which are generated by sensory inputs such as color, size and temperature; (ii) they have a geometric or topological structure that is the result of the integration of the specific 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 is not possible to attribute a value to pitch of a tone without attributing one to loudness. When quality dimensions are not integral, they are called separable.[2]

We define the notion of *domain* as a set of integral dimensions that are separable from all other dimensions. For instance, human color properties are composed of three fundamental parameters of color perception: hue, saturation and brightness (Gärdenfors, 2000, 2014). Any color perception is mapped onto some specific values to these dimensions. More generally, different colors can be described as *regions* of possible values of these three parameters (see Fig. 1).

[2] There exist several psychological tests for determining whether dimensions are integral or separable (Garner, 1974; Maddox, 1992; Melara,1992).

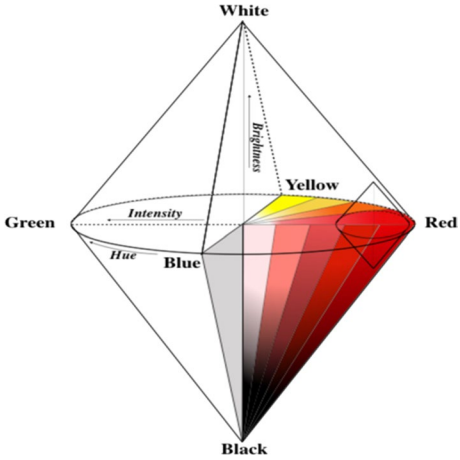

Fig. 1 Color space. The property red is represented as a convex region that corresponds to specific values of hue, saturation and brightness (color figure online)

A central criterion in this theory is that natural properties (like colors) correspond to *convex* regions of a single domain (Gärdenfors, 2000: 71). 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, the criterion assumes that the notion of betweenness is meaningful for the relevant domain.

The central notion of 'conceptual space' is defined as a collection of one or more domains with a distance function (metric) that represents properties, concepts, and their similarity relations. The distance function can vary; the most common one is the Euclidean, but also Manhattan and polar metrics may be appropriate in different contexts (see, Shepard, 1964; Johannesson, 2002; Gärdenfors, 2014).

Similarity between concepts and objects is defined a monotonically decreasing function of their distance within the space (Shepard, 1987). This makes our notion of similarity different from that of Tversky (1977), which is based on comparing the number of properties that two objects have in common with the properties where they differ.

3.2 Properties and Concepts

Many predicates in natural language, in particular those expressed by nouns, cannot be defined within a single domain, but as clusters of properties. This fact leads us to divide predicates into properties and concepts. While properties are convex regions of single domains, concepts are convex regions within a set of interconnected domains (Gärdenfors, 2000: Sec. 4.2.1). 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 crucial for inferential procedures that exploit semantic properties.

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 the Cartesian product of these five domains. And the

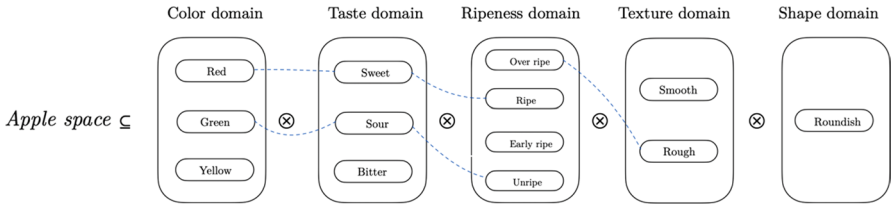

Fig. 2 The concept of an apple as a subregion of 'fruit space'. The dotted lines represent correlations between properties for the concept apple

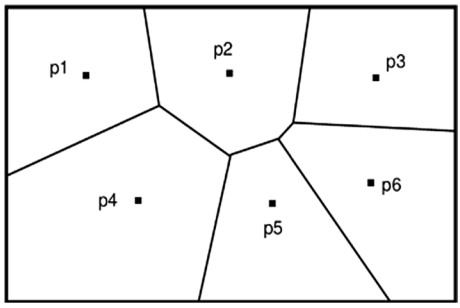

Fig. 3 Voronoi partitioning in a 2-dimensional space (color figure online)

concept apple will occupy specific subregions of these domains that correspond to the possible properties of these fruits together with correlations between regions of these domains, 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 (Gärdenfors, 2000; Lakoff, 2008; Rosch, 1975, 1983). If concepts are 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 $p_1, p_2, \ldots, p_n$ and a 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]

Within this framework, objects are seen as instances of concepts and are mapped into points of the space, and concepts are represented as regions (connected sets of points). This allows for representing graded membership and degrees of typicality (Rosch et al., 1975; Hampton, 2007); that is, we can represent objects in the space as being more or less typical instances of the categories according to their position relative to the prototype. Representing typicality in this way has several advantages

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

in terms of cognitive economy for processes like categorization (Gärdenfors, 2000, 2014) and inductive inference (Gärdenfors & Stephens, 2018; Osta-Vélez & Gärdenfors, 2020). As we will argue, this fact is crucial for representing expectations.

3.3 Context, Domain Salience, and Dynamic Conceptual Spaces

An important phenomenon that any theory of concepts must account for is that psychological similarity is a variable measure dependent on the context (Goodman, 1972). In particular, as Nosofsky (1986) noticed, conceptual similarity is modulated by attention to specific domains of the compared concepts. For instance, apples are seen (generally) as more similar to tomatoes than dates. However, in the context of choosing dessert, in which 'sweetness' is a salient feature, the similarity judgment is expected to change. Context effects have been extensively studied in the psychological literature (see Goldstone et al., 1997; Keßler et al., 2007), and geometrical models of similarity have been often criticized because of their limitations in accounting for such effects (Tversky, 1977; see Decock & Douven, 2011 for a review). The conceptual spaces model, however, doesn't suffer from these shortcomings (Johannesson, 2000; 2002). The context-sensitive character of psychological similarity is accounted for in terms of a weighted distance measure. For instance, within the context of a Euclidean metric, the distance measure will include salience weights w_i that modify the salience of dimension i in the conceptual space:

$$d(x, y) = \sqrt{\sum_i w_i(x_i - y_i)^2}$$

When a larger value is given to a weight w_i, the conceptual space is magnified along that dimension, which means that dimension i will become more important when determining the similarity between categories because larger distances (that is, dissimilarities) are penalized more when w_i is high (Gärdenfors, 2000: 20). As we will show later, this weighted distance function will have a central role for accounting for the role of context in case-based induction.

4 Typicality and Diagnosticity

We next show that conceptual spaces, thanks to their particular metric structure, allow us to model *typicality* and *diagnosticity*, which are central notions in the analysis of concepts and reasoning with concepts.

4.1 Generating Expectations

We first present how expectations can be modelled in terms of conceptual spaces. We submit that expectations depend crucially on the prototypical structure of

concepts. In this article, we only analyze expectations about the properties of the objects we reason about and do not consider relations between objects.

The underlying method of generating expectations follows from 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: 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 requires 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.

4.2 A typicality Criterion

This analysis of expectations can be formalized using conceptual spaces. The representation $\mathcal{C}(M)$ of a concept M can be seen as a subset of the Cartesian product of n domains:

$$C(M) \subseteq D_i \times D_2 \times ... \times D_n$$

An object x falling under M is represented as a n-dimensional point $x = \langle x_1, x_2, ..., x_n \rangle \in \mathcal{C}(M)$ such that each x_i in x represents the coordinates of x in the domain.

The central idea is that the expectations towards a sentence $M(x)$ are structured around the prototype p^M of M, which is also a n-dimensional point $p^M = \langle p_1^M, p_2^M, ..., p_n^M \rangle \in \mathcal{C}(M)$ with the peculiarity that each coordinate p_i^M falls within a region of D_i that represent a prototypical property of the concept M.[4] This is to say that 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, that is, to have all the prototypical properties.

Our expectations towards an object categorized as M may well go beyond the properties determined by M's prototype. For instance, if we know that x is an apple, we might expect that it is red, but we wouldn't be surprised if when we eventually see that x is actually green; after all, *green* is also a common (and as such expected) property for apples. Since most objects can have different properties in the same dimension, it makes sense to consider all non-prototypical ones as secondary expectations. In the conceptual space 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 are weaker (more defeasible) than the ones that apply to p^M. In general, we claim that for any property R in $\mathcal{C}(M)$ its *degree of*

[4] Here we assume that the prototype is a unique point. However, this may be relaxed to assuming that there is a prototypical region (Douven, 2016).

defeasibility can be specified as a function of its position relative to the prototype of M.

Our task is now to construct an ordering of properties that reflects their strength of expectations (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.[5] The criterion measures the distance of a region to a prototype via its closest point (see also Lewis & Lawry, 2016). We can use the distance function from the conceptual space to obtain this kind of information.[6] We will talk about the 'typicality degree' of a property R in $\mathcal{C}(M)$ —written "$T_M(R_i)$"— as a measure of its expectedness, and propose the following measure:

4.3 Typicality Measure

1. For any prototypical property R_i in $C(M)$, $T_M\left(R_i\right) = \min\limits_{x \in \neg R_i(x)} d\left(x, p^M\right)$
2. For any non-prototypical property R_k in a conceptual space $C(M)$, $T_M\left(R_k\right) = \min\limits_{x \in R_i(x)} d\left(x, p^M\right)$

The measure gives us a way to determine the typicality degree of every property in a conceptual space and, as a consequence, it can be used as a basis for constructing the desired expectation ordering of properties. We then define the ordering as the expectation ordering of properties associated to concept M as follows: Given two properties R_i, R_k in $\mathcal{C}(M)$, R_i is more expected than R_k (i. e., $R_i > R_k$) *iff* $T_M\left(R_i\right) > T_M\left(R_k\right)$. The typicality measure produces a fine-grained order of expectations that makes it possible to compare individual properties. Note that the degree of expectation of a property is a positive function of the strength of the nonmonotonic inference$\vdash\!\!\!\sim$. The idea of 'inference strength' can be understood in a subjective way as the 'level of confidence' that a subject has regarding some non-monotonic inference. The most entrenched properties in the ordering will generate inferences that are perceived as almost certain (e.g., "If x is a bird, then it has feathers") while less entrenched properties will produce weak and uncertain inferences (e.g., "If x is a bird, then it is small").

4.4 Diagnosticity

This notion of diagnosticity is founded on the principle that certain properties of a category can serve as key differentiators, distinguishing it from other categories within the same contrast class (see Tversky, 1977). For instance, consider the property of *having wings* in the category of BIRD. This property is highly diagnostic

[5] There are other possibilities to define the expectation ordering between properties, for example by using average distances between a prototype and a region or looking for nearest neighbors (Sadler and Shoben, 1993). It is a matter of empirical research to determine which method gives the results that best fits with how humans reason.

[6] For any point $x \in (M)$ and property $R \subseteq D_i \subseteq (M)$, "R(x)" means that the coordinate corresponding to dimension D_i falls under the subregion corresponding to R.

because it significantly differentiates birds from other contrastive animal categories that typically lack wings, such as MAMMALS or REPTILES. This kind of properties has an important role on the structure of concepts because they are maximally informative and minimize uncertainty and ambiguity during categorization.

In our model, we address the influence of the contrast class by considering the prototype of the immediate superordinate category. We can define diagnosticity in terms of typicality in conceptual spaces as follows:

1. A property R is *diagnostic* for category M if R is not a prototypical property (according to our Typicality Criterion) of the immediate superordinate category of M.
2. Given a category M, its immediate superordinate category N, and two properties R_i and R_k in$\mathcal{C}(N)$, the diagnosticity of R_i — written$D_{M,N}(R_i)$—is equal to $\frac{T_M(R_i)}{T_N(R_i)}$.

It is important to note that in the typicality and diagnosticity measures we do not count numbers of instances, but the measures are 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 properties that are probable may be atypical.[7] For example, the prototypical turtle is adult, while it is very probable that a turtle dies before adulthood.

5 Generics

The prototypical structure of concepts not only expresses itself while reasoning under uncertainty. It also becomes evident in everyday communication, particularly when we use generic sentences to convey information about the world efficiently. Generic sentences are expressions of the form "*F*s are *G*s," like "Ducks lay eggs," "The French like wine," or "Tigers are ferocious." Such sentences are central to everyday communication and cognition and they have attracted the attention of psychologists, philosophers, and linguists for a long time (e. g., Gelman, 2010; Krifka et al., 1995; Leslie & Lerner, 2016).

On the surface, generics seem to be about the prevalence of properties over groups of entities. However, when we analyze how they are used in everyday cognition, we see that statistical factors play a minor role. For instance, people usually consider the statement "Sharks kill people" valid. Yet, human deaths caused by sharks are very rare, and many species of sharks are completely harmless to humans. Likewise, the generic "Lions have manes" is often assumed to be valid even if it is only true for a relatively small subclass of ducks (mature males).

Since much of our world knowledge comes from them, generics are of great epistemological interest (Gelman, 2021). However, they are particularly difficult to analyze since they express very different kinds of information using the same linguistic format. Several attempts have been made (e.g., Cohen, 2004; Pelletier &

[7] An example of a probabilistic model is Lieto and Pozzato (2019).

Asher, 1997; Sterken, 2015b), but they all face problems (see Leslie, 2008). Some of the models are probabilistic (e.g. Tessler & Goodman, 2019; van Rooij & Schultz, 2019), but, as we have explained in the introduction, there are strong arguments against using probabilistic model to account for reasoning with concepts.

The following classical example is a reductio showing that a simple assignment of truth conditions will not work:

(1) Birds fly.
(2) Penguins do not fly.
(3) Penguins are birds.

The paradox is that all three sentences are valid generics, but they cannot all be true if they are interpreted as universal sentences.

For a logician, a typical reply is to say that (1) and (2) are not real universal sentences but implicitly mean something like "Typically, birds fly" or "Usually, penguins do not fly." Following this intuition, many have assumed that the "deep structure" of generic statements combines two predicates with a hidden operator (named "*Gen*") that refers to some adverb of quantification like "typically," "generally," or "usually." Many have believed that specifying *Gen* is the key to determining the truth-functional structure of generics, (see e.g., Pelletier & Asher, 1997), but Leslie (2008) convincingly shows that there is no satisfactory description of the semantics of *Gen*.[8] We believe that an analysis of the meaning of generics in terms of their functional role in reasoning with concepts is more fruitful than a truth-functional one.

Our proposal involves not looking upon generics as statements that can be evaluated in isolation. Instead, we take a cognitive approach that focus on their interaction with other statements that function as *expectations* in reasoning as analyzed in the previous section (cf., Leslie, 2008; Prasada et al., 2013). More specifically, our claims are that (i) generics have degrees of strength that can be evaluated according to their relation to large clusters of propositions encoding conceptual knowledge and (ii) they are particularly relevant in inductive inferences.

In the literature, there has been a discussion concerning how many types of generics should be distinguished. For example, Leslie et al. (2011) and Prasada et al. (2013) suggest that there are at least five types. We propose a new classification that distinguishes between two main kinds of generics: (a) *Property generics* dealing with characteristic properties of objects, and (b) *diagnostic generics* dealing with the diagnosticity of concepts.[9]

Our analysis builds on an elaboration of the classical distinction between 'knowledge-that' and 'knowledge-how' (Ryle, 1949) which adds '*knowledge-what*'

[8] It cannot be equivalent to any of the traditional quantifiers 'all', 'most' or 'some' (Carlson, 1977). And *Gen* cannot be systematically applied to all generics since its scope varies: "Tigers are striped" holds for most tigers while "Ducks lay eggs" only for a minority of ducks. Thus, it seems that it is impossible to describe *Gen* as a form of quantifier.

[9] This distinction is essentially the same as Pelletier and Asher's (1997) distinction between non-episodic and episodic generics, although we emphasize the role of causality.

to the classification (Gärdenfors & Stephens, 2018). Knowledge-that concerns relations between agents and propositions that describe the world, while knowledge-how concerns abilities, dispositions and actions of an agent. In contrast, 'knowledge-what' concerns the ability to *categorize*, in particular to know relations between categories and properties.

'Knowledge-what' can be broken down into three types of information about categories: Defining properties, characteristic properties and accidental facts (see also Keil & Batterman, 1984 and Gärdenfors & Stephens, 2018). These three types of information can be illustrated with an example concerning the category "spiders' web":

- *Defining* Spiders' webs are made from a protein fiber extruded from the spider's body.
- *Characteristic* Spiders' webs are used for catching insects.
- *Accidental* Spiders' webs are abundant in the cellar.

Defining properties of a category refer to information that pertains to the "core" meaning of its lexical counterpart. Characteristic properties refer to general knowledge about the category, that is, properties that generally hold of the category (exceptions may be possible). In the case when characteristic properties are formulated in sentences, the distinction between defining and characteristic corresponds to the distinction between definitional and law-like sentences that has been made within philosophy of science (Hempel, 1965). We do not assume, however, that there is a shared border between defining and characteristic properties. Accidental facts contain information about particular instances of a category.

Generics such as "Blue whales eat plankton" and "Tigers are striped" are used to express some of the characteristic properties of categories *whale* and *tiger*. A sentence such as "Tigers are mammals" is defining. In contrast, factual universals, such as "Blue whales can be seen around the Cape of Good Hope" and "Tigers can be found in the Himalayan foothills" express accidental facts about the world that are not part of the characteristic properties of the concepts.

5.1 Property Generics Express Relations Between Concepts

Property generics have the form "*A*s are *B*" or "*A*s have *B*", where *A* denotes a category (expressed by a noun or a noun phrase) and *B* denotes a property (generally expressed with the aid of an adjective). We claim that these generics do not capture any statistical fact regarding the property and the class of objects denoted by the noun, but rather express that the property is particularly relevant for the concept represented by the noun. In most cases, these generics convey information about the typicality or diagnosticity of a property. Consider for instance, the generic "Lions have manes." While most lions do not in fact have a mane (only adult males do), the generic is generally accepted because *having a mane* is both typical and diagnostic of the category LION.

Now, compare "Lions have mane" with the accidental "Lions are in the backyard." They are inferentially different: "Lions are in the backyard" is *upward entailing* (see Huang, 2011), that is, it will remain true if we substitute "lion" for any of its superordinate concepts (for example *animal*). This is clearly not the case for "Lions have mane" since "Animals have mane" is not valid.

It is well-known that generics and accidental universals behave in different ways linguistically, as pointed out by Lawler (1973):

(D) Blue whales eat plankton.
(E) A blue whale eats plankton.
(F) Blue whales can be seen around the Cape of Good Hope.
(G) A blue whale can be seen around the Cape of Good Hope.

(4) describes a characteristic property of blue whales. It can be exchanged for the indefinite singular version in (5). The generic expresses a relation between the concept blue whale and the property of feeding on plankton. In contrast, (6) is an accidental universal. A test for this is that it cannot be exchanged for the indefinite singular version in (7).

Our second claim regarding property generics is that their meaning depends on the structure of background knowledge together with pragmatic factors. In the literature, generics have been mostly analyzed as isolated expressions. We shift the perspective and see generics as expectations that must be evaluated together with other expectations (see also Leslie, 2012). We propose that property generics (as sentences expressing knowledge-what) concern compatibility relations between the semantic domain of the predicated property and the category in the sentence. In brief, generics concern more or less prototypical properties of concepts: The more prototypical a property is for a concept, the stronger will be the corresponding generic. Depending on the available information, generics might interact differently with background knowledge and be used differently in reasoning. Surprisingly enough, there has been no previous attempt in the literature to specify how our knowledge of category structure interacts with our knowledge of predicated properties.

The rationale behind this is the same as explained in Sect. 4.1: When one categorizes an object x as C, expectations about properties that x is supposed to have because of falling under C can be used to generate generics. These expectations will respect an ordering that follows from the *prototypical* structure of the concept and that it is sensitive to Grice's principles of communication. For instance, if you are informed that an object x should be categorized as a bird, but you do not know more about what kind of bird x is, then, as before, you expect that prototypical properties can be applied to x. Grice's principle of maximal informativeness says that your informant should have communicated something more specific if these generic properties do not apply to x. We, therefore, interpret a property generic "Fs are G" as "Fs that are similar to the prototype have the property G."

The Gricean pragmatic principle, plus the prototypical organization of the expectation ordering, tells us which set of generics fit better with some piece of information. In other words, it is possible to establish an ordering of generics according to their strength that will mirror the internal ordering of prototypical properties in the expectation set. For example, compare the following two valid generics:

(8) Elephants have trunks.
(9) Elephants are grey.

"Having a trunk" is a more characteristic property of elephants than being grey. One can quite easily accommodate the occurrence of a white or a black elephant, and perhaps also a pink one. An elephant, without a trunk, however, is a damaged elephant and is much less expected than a non-grey elephant. In terms of our typicality measure, a non-grey elephant is more typical than an elephant without a trunk. This is illustrated in Fig. 4, where the measures d_1 and d_2 represent the typicality of 'grey' and 'having a trunk'.

The main prediction of this approach is that the *strength* of a property generic will be a positive function of the *typicality degree* of the property within a conceptual space. The general rule is that the more typical a property is for a particular category, the stronger is the generic. This rule explains why the proportion of instances is not decisive for how useful a generic is in arguments. For example, "Books are paperbacks" is highly probable, but the generic does not concern a characteristic property.

The most important consequence of this analysis of how generics are used in reasoning is that it does not require any additional linguistic or logical operators whatsoever. In particular, no *Gen* or default operators are needed. Instead, we assume that generics can be ordered according to the strength of their expectations within the appropriate conceptual space.

5.2 The Role of Diagnosticity in Generics

Concepts can be understood as organized in hierarchical structures with a *horizontal* and a *vertical* dimension (Rosch, 1983). The horizontal dimension concerns contrast relations between concepts at the same abstraction level. For example, DOG, CAT, LION, and HORSE are in a contrasting relation since any object falling under one of them is automatically excluded from the others. This type of relation occurs when concepts are included in the same partitioning of a superordinate category (in our example, MAMMAL). The subordinate/superordinate relation is encoded in the vertical dimension. We will refer to the contrast class of concept M as the set CC(M) containing the concepts that are different from M but share its immediate superordinate category with it.

Many generics turn out to express properties that are *diagnostic* for a concept. Consider.

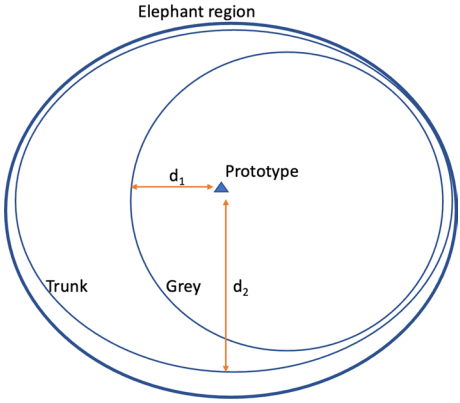

Fig. 4 The typicality of 'grey' and 'having a trunk' in the elephant region

(10) Lions have manes.

That generic expresses that.

Lions, in contrast to other felines, have manes.

In other words, having a mane is diagnostic for lions in $CC(\text{LION}) = \{\text{TIGER}, \text{CHEETAH}, \text{JAGUAR}, \text{LEOPARD}, \text{etc.}\}$.

In our model, we address the influence of the contrast class by considering the prototype of the immediate superodinate category. A property is diagnostic when it maximizes the dissimilarity between a category and the categories in its contrast class, speeding up categorization (Tversky, 1977). The measure of diagnosticity that we proposed in Sect. 4.3 assures that for any two similarly distinctive properties of a category (with respect to its contrast class), the one with the highest typicality degree will be the one with more diagnostic value.

We predict that generics with diagnostic properties are easier to endorse than generics with non-diagnostic ones because of the positive cognitive effects of learning or confirming them on agents' conceptual systems. For instance, (10) should be seen as more informative than "lions have whiskers" because 'having manes' contains information that is specific to the category and helps to differentiate it from other categories, while 'having whiskers' is common to all members in $CC(\text{LION})$ and can be inferred from characteristic knowledge of the superordinate FELINE.

Now, some properties can have diagnostic value for a category and still be non-prototypical or even rare. Consider the generic:

(K) Frenchmen eat horsemeat.

Horsemeat is atypical in the diet of most French people, although it is common in a few regions of the country; why then is (11) considered an acceptable generic? The answer is that the diagnostic value of the property in (11) in relation to *CC* (Frenchmen) compensates for its low degree of typicality. In other words, the

property *eat_horsemeat* will be more expected for Frenchmen than for any other category in *CC*(Frenchmen).

On the basis of these ideas, we propose that a generic of the form "*M* is *R*" is a valid *diagnostic generic* if *R* has high diagnostic value for *M* (with respect to a presumed superordnitae category *N*). This criterion would allow us to capture the intuition that generics with prototypical properties with diagnostic value are stronger than generics with atypical properties that also have diagnostic value: For instance, "Frenchmen speak French" is stronger than "Frenchmen eat horsemeat."

The above analysis suggests that there are two factors at work in the evaluation of a generic. The first has to do with the degree of typicality of the property in the generic, and the second with its informational contribution in the form of diagnostic value. To illustrate the difference between them, compare (10) with the generic *lions_have_bones*. Although very strong, the latter generic may seem obvious because the information in it is "inferentially available" for any competent language user. That is, minimal knowledge about the category *lion* allows you to infer that generic. On the other hand, (10) contains information that cannot be inferred from the superordinate of LION and which contributes to the specification of the category in relation to its contrast class.

In contrast to property generics, diagnostic generics do not pass Lawler's (1973) test: The generic "Frenchmen eat horsemeat" does not express the same as "Frenchman eats horsemeat" and similarly "Lions have manes" is different from "A lion has a mane". This observation supports that the two types of generics are indeed different.

In the literature (Leslie, 2008; Prasada et al., 2013; Sterken, 2015a; van Rooij, 2019), so-called "striking property generics" have been brought forward as a special type of generics. Two examples are the following:

(12) Ticks transmit Lyme disease.
(13) Sharks kill people.

Our proposal is that it is not the fact that they concern striking events that make them valid, but that they are diagnostic generics. Even though very few ticks carry the Lyme disease, this is more frequent in ticks than in any other category in CC(TICKS) (the superordinate may perhaps be BUGS–ticks are not insects but arachnids). Ticks may, in fact, be the only animals that carry the disease. Given this analysis, we reject the proposal that striking property generics form a separate class.

5.3 The Function of Generics

A fundamental pragmatic question is: What is the *use* of generics? In brief, our answer is that their main role is to express expectations of different strengths that can be used in the forms of reasoning that we present in this book. We speculate that children at an early age learn to reason with the expectations that are generated by the generics, but we know of no empirical investigations related to this position.

It seems that generics have a central role in *teaching,* in particular in what is called "natural pedagogy", that is, teaching by parents and others in everyday circumstances (Csibra & Gergely, 2009). We tell our children, already when they are small, things like: "cats say meow, dogs say woof, and cows say moo".[10] Later in school, they learn generics like "tigers have stripes", "copper conducts electricity" and "democracies have freedom of speech". Such property generics is a way of presenting characteristic properties of various categories (Leslie, 2008). Learning about categories is primarily done via their characteristic properties.[11] And when it comes to diagnostic generics containing striking properties such as "dogs bite people", they function as guidelines for caution in actions (Sterken, 2015a).

Mattos and Hinzen (2015) argue that natural pedagogy is one of the main functions of language. They write that humans have a "specific capacity to acquire, through communication, different kinds of information–respectively, *knowledge about kinds* and knowledge about particular events, actions and state of affairs which we will call here simply '*knowledge about facts*'" (Mattos & Hinzen, 2015: 7). They also argue that children learn knowledge about kinds earlier than they learn knowledge about facts.

6 Category-Based Induction

6.1 properties of Category-Based Induction

In a pioneering article, Rips (1975) studied a type of inductive inference that exploits information about individual categories for estimating the probability of property projection among them. For instance, "Dogs have sesamoid bones; thus, wolves have sesamoid bones" relies on the conceptual similarities among the categories DOG and WOLF, and not on the logical form of the argument or some other propositionally codified property. In particular, Rips saw that similarity among categories was a guiding principle for this kind of reasoning, and he proposed that the prototypical structure of natural categories also plays a role in judging the strength of inductive arguments.

Such processes, called *category-based induction* (CBI), are fundamental to our cognitive lives because of their role in dealing with uncertainty: they allow us to reason about some unknown concept *M* by exploiting information stored in our conceptual system about things that resemble *M*. They are, arguably, the clearest example of how concepts are constitutive of inductive reasoning (Feeney, 2017: 167). Understanding how CBI works, and especially which features of our conceptual systems this form of reasoning exploits, can shed light on the general problem of the role of concepts in inferences. In this section, we discuss the general features of CBI and show how conceptual spaces can model them.

[10] Children's picture books of animals and other object categories highlight the diagnostic properties of the categories.

[11] Van Rooij (2018) emphasizes the role of learning generics.

CBI arguments are composed of generic sentences (e.g., 'Dogs have sesamoid bones,' or 'Bears love onions') both in the premise(s) and in the conclusion. We abbreviate an inference of the form '*X* have property *R*; thus, *Y* have property *R*' as '$X \rightarrow Y$'. One argument for this abbreviation is that, in almost all studies, subjects typically have little or no knowledge about the property *R* and therefore it does not influence the strength of the argument.[12]

CBI arguments can be classified in two major ways: according to their number of premises and whether the conclusion is at the same conceptual level as the premises or in some superordinate category. When the premise(s) and conclusion categories are at the same conceptual level the argument is called 'specific,' as in ROBIN → CROW. When the argument involves a generalization (a "jump" to a superordinate conceptual level), then it's called 'general,' as in TABLE → FURNITURE.

The empirical literature has shown that the most robust criterion used in CBI is similarity among categories (Rips, 1975; Carey, 1985; Osherson et al., 1990; Lopez et al., 1992). This can be formulated as that our expectations regarding property projection among two categories *X* and *Y* are a positive function of the similarity between *X* and *Y*. For instance, arguments like "Ostriches are *R*, then emus are *R*" are generally seen as stronger than arguments like "Ostriches are *R*, then blue jays are *R*" simply because ostriches are more similar to emus than to blue jays.

The typicality of the categories in the premises of CBI arguments also have a positive effect on the expectations of property projection. For instance, the inference "Robins have enzyme E; thus, ostriches have enzyme E" is often judged as stronger than "Penguins have enzyme E; thus, ostriches have enzyme E", because the category ROBIN is more typical of BIRD than the category PENGUIN. Hampton and Cannon (2004) have shown that arguments with a highly typical category in the conclusion (like CHICKEN → ROBIN) are judged as stronger than arguments with non-typical conclusion categories (like CHICKEN → VULTURE).

This typicality effect also produces what is called 'asymmetry,' that is, the fact that switching the categories from the premises and conclusion often changes the expectations of property projection, according to the degree of typicality of the category in the premise(s). For instance, arguments like "Cows have enzyme E; thus, otters have enzyme E" is considered stronger than arguments like "Otters have enzyme E; thus, cows have enzyme E" since cows are more typical mammals than otters.

Another important aspect is that subjects assume a common superordinate category of the premises when making inferences or judging the strength of this kind of argument. Sometimes this superordinate category appears explicitly in the conclusion; other times, it is just considered implicitly. Four important phenomena related to such evoked superordinate category have been studied in the empirical literature: homogeneity, monotonicity, nonmonotonicity, and premise diversity.

Homogeneity refers to the idea that the more abstract and less homogenous the category in the conclusion is, the weaker the argument. For instance, arguments like

[12] This is the standard procedure in most CBI studies. Of course, it would be interesting to investigate the influence of an R explicitly, but that would make the experimental procedure much more cumbersome.

"Robins are S and blue jays are *R*; thus, all birds are *R*" are judged stronger than "Robins are *R* and blue jays are *R*; thus, all animals are *R*." This is not surprising at all. As we said before, we deal with different degrees of uncertainty when evaluating arguments or making inferences that involve generalizations. The more abstract the category in the conclusion, the more information we need from the premises to cover it.[13]

A possible way of explaining this is by referring to similarity and typicality as the two main criteria for using categories. Basic level categories are more homogenous. As such, it is easier for us to apply criteria of similarity among their members. Abstract categories are more diverse and less homogenous, so comparing their members in terms of similarity is more complex (for instance, the category ANIMAL include highly dissimilar subcategories, such as ELEPHANT and STARFISH). Along the same line, basic categories have clear prototypes, while it is complicated for us to construct prototypes for abstract categories (see, Ungerer and Schmid, 2006: Ch. 2 for an explanation). In this sense, typicality, considered as a criterion for using categories, is stronger in basic-level categories than in abstract ones.

Monotonicity refers to the fact that the addition of premises, as long as their categories are included in the evoked superordinate category, strengthen the argument (Osherson et al., 1990). For instance, an argument of the form (ROBIN & HAWK) ⟶ BIRD is weaker that an argument of the form (ROBIN & HAWK & PIGEON) ⟶ BIRD. However, if we add to the premises a category that is not from the evoked superordinate category, then the argument becomes weaker. This is called 'nonmonotonicity.' For instance, an argument with the categories (PEACOCK & CROW) ⟶ BIRD is stronger than an argument that goes from (PEACOCK & CROW & RABBIT) ⟶ BIRD.

Finally, empirical studies have shown a 'diversity effect' in CBI (Feeney & Heit, 2011; Osherson et al., 1990): Arguments like "Horses have an ulnar artery and seals have an ulnar artery; thus, all mammals have an ulnar artery" are considered as stronger than the argument "Horses have an ulnar artery, and cows have an ulnar artery; thus all mammals have an ulnar artery." The less similar the categories in the premises are, the stronger the argument tends to be. An interesting way of understanding this phenomenon builds on the idea of 'category coverage' (Osherson et al., 1990). As we mentioned before, when performing or evaluating categorical inductions, we take as a reference (implicitly or explicitly, according to whether we deal with a specific or general argument) some superordinate category that includes all the categories in the premises. The strength of the argument will depend, to some extent, upon how the categories in the premises *cover* this superordinate category. For instance, similar categories like HORSE and COW have less coverage of the superordinate category than dissimilar categories like HORSE and SEAL. In this sense, coverage can be described in terms of similarity.

[13] As argued by Sloman and Lagnado (2005:106), we seem to have a 'preferred level of induction' that coincides with what Mervis and Rosch (1981) called 'basic-level' categories, i.e., categories with an intermediate level of specificity (e.g., DOG or CHAIR).

6.2 Modeling CBI in Conceptual Spaces

Osta-Vélez & Gärdenfors, (2020) introduced a model of CBI based on conceptual spaces. In our modeling, we use the expression '$ExpR(X \rightarrow Y)_Z$' to stand for *the expectation that the property R is projected from category X to category Y,* with *Z* as the lowest-level superordinate category that contains both *X* and *Y*. Here we focus on the simplest case of category-based inference: single premises/specific arguments.[14] For this kind of inductive inference, we want $ExpR(X \rightarrow Y)_Z$ to satisfy the following criteria:

- It is positively correlated with $sim(X, Y)$, where $sim(X, Y)$ denotes the similarity of *X* and *Y*;
- It is positively correlated with $sim(X, p^Z)$, where p^Z is the prototype of *Z*;
- It is positively correlated with $sim(Y, p^Z)$.

The rationale for the first condition is that the more similar the categories *X* and *Y* are, the more expected will it be that *Y* has property *R* if *X* has it. Regarding condition (ii), the intuition is that the more prototypical *X* is, the more expected it is that another category *Y* has property *R*, given that *X* has it. Condition (iii) is motivated by Hampton and Cannon's (2004) conclusion-typicality: the more prototypical *Y* is the more expected it is that *Y* has property *S* if *X* has it.

To illustrate the basic idea of our approach with a simple case, let us assume that *X* and *Y* are small regions so that we can identify them with points in a conceptual space.[15] Then, given a conceptual space representing the categories *X*, *Y*, and *Z* and the distance function *d* of the space, we can account for the three conditions above by the following equation:

$$ExpR(X \rightarrow Y)_Z = (d(X, Y) \cdot d(X, p^Z)^a \cdot d(Y, p^Z)^b)^{-1} \tag{1}$$

where *a* and *b* are positive constants such that $a > b$. This assumption expresses that premise typicality contributes more to expectations than conclusion-typicality since, according to the literature, the former is a more prevalent phenomenon than the latter. The values of both *a* and *b* must be empirically determined from data about CBI judgements.

Now, following Shepard's (1987) universal law of generalization, which claims that similarity is an exponentially decreasing function of distance, we can take the logarithm of (i) and obtain:

$$logExpR(X \rightarrow Y)_Z = sim(X, Y) + a \cdot sim\left(X, p^Z\right) + b \cdot sim\left(Y, p^Z\right) \tag{2}$$

By convention, for any two categories *X* and *Y*, $0 \leq sim(X, Y) \leq 1$ and $sim(X, Y) = 1$ *if* $X = Y$.

Now, equation (2) captures the basic idea that for single-premise specific arguments the expectations of property projection among categories are

[14] For a more general analysis, see Osta-Vélez, M., & Gärdenfors, P. (2020).

[15] For a more general treatment, see Osta-Vélez, M., & Gärdenfors, P. (2020).

determined by a weighted sum of three factors: premise-conclusion similarity, premise-typicality, and conclusion-typicality.

Equation (1), applied to a set of prototypes for categories, captures similarity, premise and conclusion typicality and asymmetry effects in CBI. For instance, when considering the sentence "emus have property R," people expect more that ostriches also have *property R* than that penguins have it. This is due to the similarity effect since $sim(\text{emu}, \text{ostrich}) > sim(\text{emu}, \text{penguin})$. If we construct a "bird space" through some set of prototypes, this inequality would be immediately represented by the relative positions in the space of the two pairs $<$ EMU, OSTRICH $>$, and $<$ EMU, PENGUIN $>$ (see Fig. 5). And it can be measured via the distance function of the space. Since $sim(\text{emu} \rightarrow \text{ostrich}) > sim(\text{emu} \rightarrow \text{penguin})$, it follows that $ExpR(\text{emu} \rightarrow \text{ostrich})_{\text{bird}} > ExpR(\text{emu} \rightarrow \text{penguin})_{\text{bird}}$.

As we mentioned, this model also predicts asymmetry and premise and conclusion-typicality. For instance, $ExpR(\text{robin} \rightarrow \text{emu})_{\text{bird}} > \text{Exp}R(\text{emu} \rightarrow \text{robin})_{\text{bird}}$ since $sim(\text{robin}, p^{\text{bird}}) > sim(\text{emu}, p^{\text{bird}})$ and $a > b$. Regarding conclusion-typicality assume, following the bird space in Fig. 4, that $sim(\text{ostrich}, \text{vulture}) \approx sim(\text{ostrich}, \text{robin})$ and that $sim(\text{ostrich}, p^{\text{bird}}) \approx sim(\text{vulture}, p^{\text{bird}})$. Then $ExpR(\text{ostrich} \rightarrow \text{robin})_{\text{bird}} > \text{Exp}R(\text{ostrich} \rightarrow \text{vulture})_{\text{bird}}$ since $sim\left(\text{robin}, p^{\text{bird}}\right)$ is significantly larger than $sim\left(\text{vulture}, p^{\text{bird}}\right)$.

The model presented so far only concerns arguments with blank properties. However, there is robust evidence that the perceived strength of CBI arguments can be influenced by knowledge that agents have about the properties in the arguments as well as of possible causal relations between properties and categories (see Rehder et al., 2001; Coley et al., 2005). For reasons of space, we will not elaborate on this problem here. However, in Osta-Vélez, & Gärdenfors, (2020) we indicate possible ways of modeling this through manipulation of the weights of certain dimensions in the distance functions of the space.

6.3 Experimental Evidence

There is a wealth of experimental evidence concerning category-based induction (see Heit, 2000; Feeney, 2017). Osta-Vélez, & Gärdenfors, (2020) show that our model can account for a large majority of the results of these experiments. Most of the experiments, however, only report qualitative results. Our model also allows quantitative predictions once the metric of underlying conceptual space has been estimated. Douven et al. (2021) present a study that explicitly tests some such predictions of our model. Rather than descriptions of objects, the stimuli used were generated from 49 pictures of objects from Douven (2016) that look like cups, vases and bowls and which are organized 7-by-7 along the dimensions of width and height. The psychological distances between the objects have been carefully determined in Douven (2016). Consequently, a distance function for the space of the objects was already available.

The experiment tested two hypotheses generated from our model: (i) The strength of proximity-influenced arguments can be predicted from the distance in the conceptual space between the object in the premise and the object in

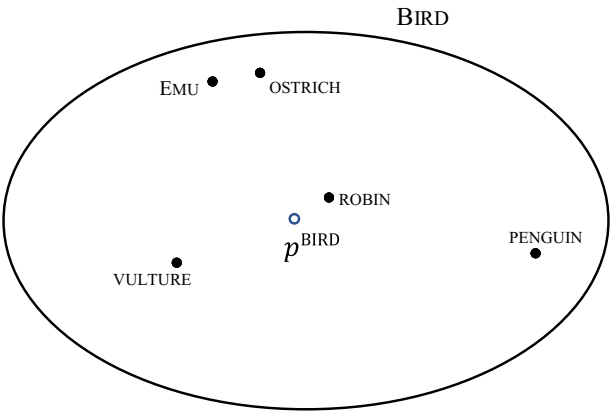

Fig. 5 "Bird space" representing the positions of the different bird categories relative to a prototype

the conclusion, as well as from how typical these objects are of the projected property. (ii) Truth ratings of conditional sentences mentioning two embodying proximity-influenced inference can be predicted on the exact same basis.

To test the first hypothesis, the participants were shown 12 pairs of objects randomly selected from the 49 in the space. For each pair, the participant was asked to suppose that the vessel that appeared on the left was a vase and then to indicate whether that gave them reason to believe that the vessel on the right was a vase as well. The response had to be given on a 7-point Likert-scale. For the second hypothesis, the participants were shown the same 12 pairs. In this part, the participants were asked to evaluate the truth of a conditional sentence of the form "If the vessel on the left is a vase, so is the vessel on the right." They were then asked to indicate whether they thought this was true or false, but they were also given the option "neither".

The results provided strong support for both hypotheses: antecedent–consequent similarity was strongly associated with perceived argument strength, respectively, degree of truth. For further details, see Douven et al. (2021).

7 Analogies

So far, we have analyzed inferential mechanisms that use similarity relatively straightforwardly; but there are even more sophisticated ways our cognition exploits this relation. Analogy is a paradigmatic example of this. For instance, consider the sentence.

(N) Cheetahs are like race cars.

If (14) is read as expressing a straightforward "overall" similarity between the categories (as in "leopards are like cheetahs"), then it would not make sense to any

competent language user. However, (14) is meaningful under an interpretation in which the similarity relation is understood as focusing on a salient feature shared between the categories (speed); in other words, if it is read as an analogy.

From a semantic and cognitive perspective, analogical statements are rather peculiar: Their role is not to convey information about states of affairs in the world but to enrich and structure our conceptual knowledge by pointing out similarity relations across seemingly distant fields of knowledge. In this sense, they have an epistemic function that appears particularly relevant to the organization and formation of abstract categories (see Gentner & Hoyos, 2017).

While the literature on analogies is vast, there are no general frameworks that explain analogical mechanisms as part of the family of cognitive processes that exploit conceptual similarity. This section will show how conceptual spaces can address this gap. In particular, we will argue that analogy is a domain-specific mechanism that depends on dimensional salience and that can be characterized as a search procedure.

Two types of analogical structures have monopolized the attention in the last decades: *direct* and *composed* analogies. The former compares an individual source with an individual target, like in (14). The latter compares two pairs of objects or categories according to some salient relation between the elements in each pair. For instance, the sentence "the foot is to the leg as the hand is to the arm" is a composed analogy since the salient (mereological) relation between *foot* and *leg* is *symmetrical* to the mereological relation between *hand* and *arm*. In this section, however, we focus on composed analogies. We will use the notation *A:B:C:D* for a composed analogy where the pair *A:B* is compared to the pair *C:D*.

7.1 Dimensional Salience

The approach advanced here builds on two observations. First, we claim that composed analogies need to be analyzed in the light of a theory of conceptual structure. Second, we propose that in most cases, analogical similarity depends on dimensional salience, more precisely, on identifying one or more dimensions (domains) that will serve as a frame of comparison for the categories in the analogy. The degree of salience of these dimensions for the given categories correlates with the analogy's 'quality' or 'aptness.' This last idea is rather straightforward. Consider the following analogies:

(O) Dog:puppy::cat:kitten
(P) Sweet:apple::sour:lemon
(Q) Hot:warm::cold:cool
(R) Rabbit:lion::tuna:shark

Each of these analogies consists of projecting a salient semantic relation among the categories of the first pair into the categories of the second pair. This relation depends on identifying one or more dimensions of the categories that can serve

as 'analogy factors'. In (15), the analogy factor is the *age* dimension, in (1) it is the *taste* domain, in (17) the *temperature dimension,* and in (18) *size* and *ferocity* dimensions. The analogy factor is generally differential: it selects a dimension in which the categories of the first pair have significantly different values. A challenge while evaluating an analogical relation is to identify, among the many dimensions that can constitute the categories involved, which are the ones that can better bear the analogical relation. For example, *size* can be a good candidate for analogy factor in (1), but *color* clearly not. In our approach, the dimensions that are going to have priority as potential analogy factors are the most *salient* dimensions of the categories in the first pair. Such a salience factor is difficult to model in proposition-based computational implementations.

A straightforward prediction of this approach is that the processing speed of an analogy will be positively correlated with the degree of saliency of the analogy factors and negatively correlated with the number of dimensions that can be considered as potential analogy factors. For instance, (17) is a straightforward analogy because a unique dimension relates its four categories; (18), on the other hand, offers multiple possible dimensions as potential analogy factors and, as a consequence, it has a higher degree of analogical complexity.While classical approaches tend to look for highly general models of analogy (Gentner, 1983; Holyoak & Thagard, 1989), our view departs from the idea that analogy is concept-specific. Our position is that analogies exploit properties of the representational structures associated with the words that appear in them. Since different word classes represent different kinds of concepts (Gärdenfors, 2014), we need a theory that integrates different sub-models. In contrast to traditional approaches in logic and computer science where all predicates are treated on a par, we aim to show that dividing them into their different conceptual roles will yield more fruitful computational systems, specified as different search procedures.

7.2 Analogy as Search

From a computational perspective, we propose that analogy can be understood as a *search procedure*, that is, as a problem-solving strategy that consists of searching on a database for an element that meets a specific condition established by the problem. We then characterize analogical problems as having the following components:

Search space A set of concepts in a *lexicon L.*
Initial state A:B::C:X (with *X* unknown).
Goal condition Find (at least) one element *X* in *L* such that the semantic relation in *A:B* is replicated in *C:X.*
Search algorithm To be defined after an analysis of the kind of semantic relation in the initial state.
Final state A concept (or preference order of concepts) satisfying the goal condition.

In this framework, the process of solving (or verifying) an analogy begins by identifying the semantic relation in *A:B* and restricting the search space accordingly in the light of *C*. For instance, the analogy *red:apple::yellow:X* is about a fruit category and a prototypical property in the color dimension. The search space for *X* will be the conceptual space of *fruit* and the goal condition will be satisfied by fruit categories for which the color yellow is prototypical.[16]The second step consists of specifying the search algorithm, which will depend on the semantic relation to be replicated. Once decided its type, the algorithm is applied to the restricted search space looking for one or more categories that satisfy the goal condition: to find an element that complete a 'semantic symmetry' between *A:B* and *C:X*.

An important point is that this semantic relation can be of various types. We distinguish between (i) categorical (dimensional) relations (e.g., *tuna:shark* or *hot:cold*), (ii) property-category relations (e.g., *yellow:lemon*), (iii) event-based relations (e.g., *open door:closed door*), and (iv) part-whole relations (e.g., *foot:leg*). In Osta-Vélez, & Gärdenfors, (2022b), we advanced several search algorithms based on conceptual spaces that model the reasoning underlying these analogies-types. In this section we will limit ourselves to describing two of these algorithms.

7.3 Category-Based Analogies

One of the earliest models of analogy was developed by Rumelhart and Abrahamson (1973), who showed that it is possible to express analogical similarity as a function of the semantic distance between categories represented as points in a multidimensional space. In particular, they claimed that analogies of the form *A:B::C:D* must follow a 'parallelogram rule' according to which the vectorial distance between categories *A* and *B* must be equal (or highly similar) to the vectorial distance between *C* and *D*.

In a series of experiments using Henley's (1969) 3-dimensional mammal-space (see Fig. 6 for some examples), Rumelhart and Abrahamson showed that when presented with analogy problems like *monkey:pig::gorilla:X*, with *rabbit*, *tiger, cow*, and *elephant* as alternatives for *X*, subjects rank the four options following the parallelogram rule. The parallelogram model predicts that *cow* is the preferred solution. Their experiment clearly supported the model.

We propose a generalized version of the parallelogram model which follows the semi-algorithmic approach described in the introduction. The basic idea is that the conceptual space in which the vectorial comparison is carried out is not fixed, but rather depends on the dimensions that are taken as analogy factors in each specific analogy.

In our model, the categories in a category-based analogy *A:B::C:D* are convex regions of a common conceptual space *M*, since they are all at the same conceptual level. For the sake of simplicity, we assume that each of these categories has a precise prototype represented by a point in the space. For category *X,* we refer to that point as p^X. The following describes the main steps of the search algorithm.

[16] Another type of analogy involves finding a new domain where the semantic relation between A and B corresponds to the relation between c and X. For example, "The immune system is to the body like police to a society" which results in some understanding of an immune system. We are grateful to an anonymous reviewer for pointing this out.

(i) Given a composed analogy *A:B::C:X*, with *X* unknown, the first step in the process consists in finding the smallest conceptual space *M* such that *A*, *B*, $C \subset M$. This space corresponds to the immediate superordinate category of *A*, *B,* and *C*. For instance, in *tiger:rabbit::eagle:X*, *M* will be *animal* but, in *tiger:rabbit::truck:X, M* will be *thing*. *M* will be the *search space* in which the algorithm will operate. Notice that the number of dimensions apt for establishing an analogical comparison depends on the specificity of *M* (that is, its place in Rosch's (1983) vertical level of categorization). Dimensions that are available for animals in *tiger:rabbit::eagle:X* like *diet*, *ferocity*, or *humanness* cannot be applied to *things* in *tiger:rabbit::truck:X*.

(ii) The second step consists in selecting from *M* a set of salient dimensions D_1, D_2,…,D_n where the salience is generally determined by the difference between *A* and *B* (often only one dimension is relevant). For instance, consider the relation *tiger:rabbit::eagle:robin*. *Ferocity* and *size* are two salient dimensions of *animal*, since an important difference can be established between *tiger* and *rabbit* across these dimensions. If these differences can be replicated for categories *eagle* and *robin*, then the analogy is sound. The choice of these dimensions as frame of comparison will generate a 'new' lower-dimensional conceptual space M^* with a distance function d^*.[17] This modulated distance function will be used to compute what we have called *analogical similarity* and constitutes the main difference with the Rumelhart and Abrahamson's parallelogram model.

(iii) The last step of the search algorithm is the application of the parallelogram rule on M^* for choosing the optimal solution to *X* in *A:B::C:X*. For this, we start with the prototypes p^A, p^B, p^C, and the vector $\overrightarrow{p^A p^B}$ in M^*, and we find the point $y \epsilon M^*$ that is the head of a new vector $\overrightarrow{p^C y}$ that is as close as possible (same direction and magnitude) to $\overrightarrow{p^A p^B}$. The category $X \subseteq M$ that gives the strongest analogical relation will be the one whose prototype p^X is closer to *y* than any other prototype in *M*, that is, p^X such that $d^*(p^X, y) < d^*(p^Z, y) \forall p^Z \in M$.

Let us illustrate this procedure with a toy example. Consider the incomplete analogy *mouse:wolf::rabbit:X* and a reduced search space with categories *hippo, buffalo, elephant,* and *gorilla*. *M* will be the mammal space used by Rumelhart and Abrahamson (1973) (see Fig. 6), and the dimensions that will serve as frame of comparison in the space *M** will be *size* and *ferocity*, due to the salient difference that the categories *mouse* and *wolf* maintain across them. The *humanness* dimension in Fig. 6 is less salient and will not be part of *M**. Then, in a weighted conceptual space *M**, a point *y* will be determined as the head of a vector with tail in the prototype of *rabbit* that is equivalent to the vector formed by the prototypes of *mouse* and *wolf*. Assuming the positions of the prototypes as depicted in Fig. 7, the prototype of *buffalo* is the optimal solution to the analogy since it is closer to *y* than any other prototype in *M**.

Analogies are not all or nothing, but have degrees of aptness or soundness. For instance, categories that are very close (in the weighted conceptual space) to the

[17] Notice that *M* and *M** are the same search space since they include the same set of subcategories.

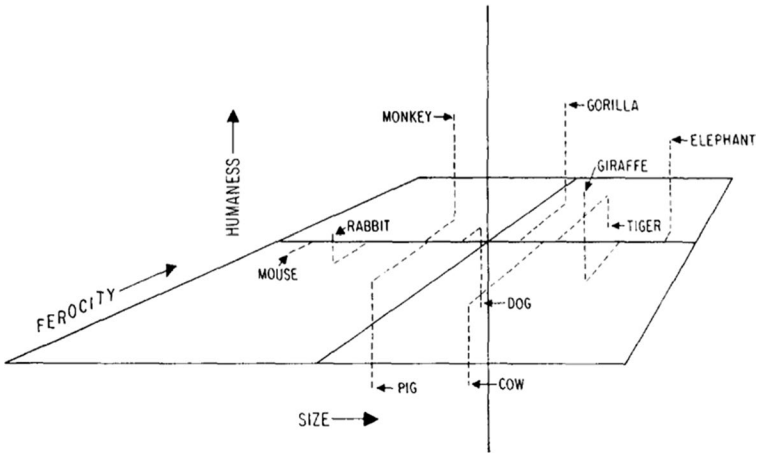

Fig. 6 Mammal space organized around the size, humanness, and ferocity dimensions. From Rumelhart and Abrahamson (1973: 3)

optimal choice in a category-based analogy might also be good solutions. In addition to this, it is possible that different sets of dimensions are taken as analogy factors, generating multiple possible sound analogies. We believe that, for most cases, there is a particularly salient set of dimensions that will produce the strongest analogical relation. However, offering a systematic criterion for finding it is rather complicated because it is strongly dependent on the subjects' knowledge of a particular semantic domain, as well as on the semantic intuitions rooted in a community of speakers. Ultimately, finding the set of salient dimensions for a given category is an empirical question.[18]

7.4 Property-Based Analogies

An important limitation of Rumelhart and Abrahamson's (1973) model is that it can only deal with analogies at the same conceptual level. Consider the following two examples:

(19) *Apple:red::banana:yellow.*
(20) *Fish:swim::bird:fly.*

(19) and (20) are sound analogies, but they cannot be analyzed in terms of the parallelogram model. How can we compare a color with a fruit or an animal with a means of motion? From a formal perspective, there is no way of comparing two vectors from different conceptual spaces.

We call analogies like (19) and (20) *property-based analogies*. Naturally, since the semantic relation between the pairs of terms in these analogies differs from that

[18] Some empirical methods for determining dimensional salience in natural categories can be found in Sloman et al. (1998) and Rein et al. (2007).

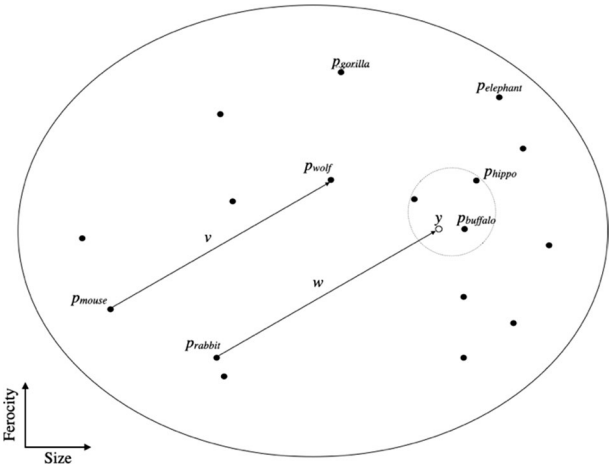

Fig. 7 Mammal space M^* with weights in the size and ferocity dimensions (based on Rumelhart & Abrahamson, 1973)

characterizing category-based analogies, explicating them via a search algorithm requires a different approach. In particular, the search space for X will be the set of lexical items associated to common properties of the category in the C-part of the analogy.

Our proposal for property-based analogies is rather straightforward: We claim that the strength of an analogy depends on two factors: first, on the identification of the dimension(s) that corresponds to the property in the pair and second, on identifying the *typicality degree* of that property for the category in the pair. In other words, we evaluate the aptness of these analogies by checking that the properties in the pairs are from the same dimension and, with the aid of our typicality condition, that they are similarly expected for the category in the pair. In this sense, an analogy like (20) must be considered as stronger than the variant *fish:swim::bird:walk* because, even if birds can walk, flying is more typical than walking for that category (see Osta-Vélez, & Gärdenfors, 2022b).

Given an analogy $A{:}B{::}C{:}X$ where A and C are categories and B is a property of A in dimension D, the choice for X which gives the strongest analogical relations is a different property in D whose typicality degree is closer to B's typicality degree than the typicality degree of any other property in D. We predict that if there are various properties in D with the same typicality degree as B for category C, then the analogy will be weaker than for categories for which this is not the case. For example, the analogy *lion:beige::raven:black* must be judged as stronger than *lion:beige::dog:brown* because several colors other than *brown* are also typical for the category *dog*.

An important contribution of our approach is its detailed analysis of the role of semantic similarity in analogy. Propositional-based views, such as Gentner's (1983), or Thagard's (Thagard et al., 1990) also rely on semantic similarity but ignore the modulating role of dimensions and salience. We show that focusing on conceptual

structure rather than propositional structure has clear advantages for explaining the diversity of analogies and for designing modeling algorithms.

8 Conclusion

Within philosophy, reasoning with concepts has received comparatively little attention since such reasoning does not conform to the assumption that the validity of an argument depends only on the formal structure of premises and conclusion. In contrast, research concerning concept-based reasoning has been a lively field within psychology, generating a wealth of experimental investigations. However, this research has been developed in a rather fragmentary manner, using different theoretical frameworks and modeling tools. Clarifying the relationship between reasoning and concepts has thus become necessary. Given the intricate nature of this relationship and the diverse mechanisms through which it unfolds, it would be beneficial to approach this task within the confines of a unified theory.

As a proposal for unifying these areas and building on the idea that rational inference exploits properties of conceptual structure rather that syntactic properties of language, we have presented a framework based on conceptual spaces. Using the distance measure provided by the conceptual spaces, we can model *typicality* as distances to prototypes and *similarity* as distances between points. These measures play a central role for all the types of reasoning we have considered. By exploiting the measures, our model thereby allows for new quantitative predictions, which cannot be made from previous psychological models. Here, we have focused on expectations, generics, category-based induction and composed analogies, but the general framework of conceptual space can also be applied to other related areas, such as direct analogies, metaphors, and causal reasoning. The model also makes it possible to compare the predictions and the results from the different subareas, as is the case of the relationship between generics and non-monotonic reasoning based on expectations.

We claim that our model offers a significant advantage over earlier approaches as it provides greater explanatory depth. While other methods focus solely on some computational aspects of inference and assume a particular format of mental representation of conceptual information without critical evaluation, our model explains both why concepts have the structure they do and how inferential mechanisms operate over these structures. This level of detail provides a more comprehensive understanding of the mechanisms that underlie human reasoning and decision making.

Our approach can be empirically tested through a number of well-established experimental paradigms in psychology. These include Induction Tasks, which examine participants' ability to make generalizations based on specific instances or categories (e.g., Heit, 1998); Feature Listing, which involves eliciting attributes of a concept from participants, with frequency and order of features offering insight into their conceptual understanding (e.g., Sloman et al., 1998); and Exemplar ratings and typicality judgments, which involve categorizing and rating objects based on

their resemblance to previous examples or a prototypical instance (e.g., Verheyen & Égré, 2018). Furthermore, Spatial Arrangement tasks ask participants to physically arrange items based on perceived similarity, offering a spatial representation of conceptual relationships (see Richie et al., 2020), and Multidimensional Scaling (MDS) uses distance-based statistical techniques on sets of similarity judgments to visually represent the structure of conceptual spaces, allowing for a robust analysis of conceptual similarity and dissimilarity (e.g., Rips, 1975). Collectively, these diverse paradigms provide a rigorous empirical framework to test the validity and applicability of the models presented herein.

In this review, we have not directly considered computational applications of the model. However, the possibility of developing such applications has been one of our motivations for presenting a model that is based on distance measures in conceptual spaces. Since there already exist computational models of conceptual spaces (Adams & Raubal, 2009; Chella et al., 2001; Gärdenfors, 2014; Lieto, 2021, Wheeler et al., 2022), our accounts of the different forms of concept-based reasoning can, in principle, be computationally implemented. The first step would be do describe *domain structures*. This involves, above all, specifying their geometric and topological structure. The second step is to give information about how the resulting space is partitioned into concepts. Using prototypes and Voronoi tessellations, this can be done in a computationally efficient way. Representing information by conceptual spaces requires computations that involve vectors, using inferences based on similarities, rather than mechanisms based on tree searching in a rule-based symbolic approach. Developing such implementations of conceptual spaces could enable the creation of new forms of automated reasoning that go beyond systems based on logical formalisms or neural networks, something that would be useful in the efforts to simulate common-sense reasoning with concepts.

Funding Open Access funding enabled and organized by Projekt DEAL. Matías Osta-Vélez was supported by the Deutsche Forschungsgemeinschaft (DFG) Grant SCHU 566/17.1 “Parameterised Frames and Conceptual Spaces”.

References

Adams, B., & Raubal, M. (2009). Conceptual space markup language (CSML): Towards the cognitive semantic web. In Third IEEE International Conference on Semantic Computing (ICSC 2009), Berkeley, CA, 253–260.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(4), 629–654.
Bartha, P. (2010). *By parallel reasoning*. Oxford University Press.
Bonatti, L. (1994). Why should we abandon the mental logic hypothesis? *Cognition, 50*(1–3), 17–39.
Brewka, G. (1991). *Nonmonotonic reasoning: logical foundations of commonsense* (Vol. 12). Cambridge University Press.
Carey, S. (1985). Are children fundamentally different kinds of thinkers and learners than adults. *Thinking and Learning Skills, 2*, 485–517.
Carlson, G. (1977). Reference to Kinds in English. PhD dissertation, University of Massachusetts, Amherst
Chella, A., Frixione, M., & Gaglio, S. (2001). Conceptual spaces for computer vision representations. *Artificial Intelligence Review, 16*(2), 137–152. https://doi.org/10.1023/A:1011658027344
Cohen, A. (2004). Generics and mental representations. *Linguistics and Philosophy, 27*, 529–556.
Coley, J., Shafto, P., Stepanova, O., & Baraff, E. (2005). Knowledge and Category-Based Induction. In W.-K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin* (pp. 69–85). American Psychological Association.
Connolly, A. C., Fodor, J. A., Gleitman, L. R., & Gleitman, H. (2007). Why stereotypes don't even make good defaults. *Cognition, 103*(1), 1–22.
Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153.
Decock, L., & Douven, I. (2011). Similarity after goodman. *Review of Philosophy and Psychology, 2*(1), 61–75.
Douven, I. (2016). Vagueness, graded membership, and conceptual spaces. *Cognition, 151*, 80–95.
Douven, I., Verheyen, S., Gärdenfors, P., Elqayam, S., Osta-Vélez, M. (2021). Similarity-based reasoning in conceptual spaces. Submitted.
Feeney, A. (2017). Forty years of progress on category-based inductive reasoning. In L. J. Ball & V. A. Thompson (Eds.), The Routledge international handbook of thinking and reasoning (pp. 167–185). Routledge/Taylor & Francis Group.
Feeney, A., & Heit, E. (2011). Properties of the diversity effect in category-based inductive reasoning. *Thinking & Reasoning, 17*(2), 156–181.
Fodor, J. A. (1983). *Representations: Philosophical essays on the foundations of cognitive science*. MITit Press.
Fodor, J. A., & Pylyshyn, Z. W. (2015). *Minds without meanings: An essay on the content of concepts*. MIT Press.
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.
Garner, W. R. (1974). *The Processing of Information and Structure*. Erlbaum.
Gelman, S. A. (2021). Generics in society. *Language in Society, 50*(4), 517–532.
Gelman, S. A. (2010). Generics as a window onto young children's concepts. In F. J. Pelletier (Ed.) *Kinds, things, and stuff: Mass terms and generics* (pp. 100–121).*Oxford University Press.*
Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*(2), 155–170.
Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in Cognitive Science, 9*(3), 672–693.
Goldstone, R. L., Medin, D. L., & Halberstadt, J. (1997). Similarity in context. *Memory & Cognition, 25*(2), 237–255.
Goodman, N. (1972). Seven strictures on similarity. *Problems and Projects* (pp. 437–446). Bobbs-Merrill.
Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics* (Vol. 3, pp. 41–58). Academic Press.
Hampton, J. A. (2007). Typicality, graded membership, and vagueness. *Cognitive Science, 31*(3), 355–384.

Hampton, J. A., & Cannon, I. (2004). Category-based induction: An effect of conclusion typicality. *Memory & Cognition, 32*(2), 235–243.

Hampton, J. A., Passanisi, A., & Jönsson, M. L. (2011). The modifier effect and property mutability. *Journal of Memory and Language, 64*(3), 233–248.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review, 7*(4), 569–592.

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.). *Rational models of cognition* (pp. 248–274). Oxford University Press.

Hempel, C. (1965). *Aspects of scientific explanation*. The Free Press.

Henley, N. M. (1969). A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior, 8*(2), 176–184.

Hofstadter, D. R., & Sander, E. (2013). *Surfaces and essences: Analogy as the fuel and fire of thinking*. Basic books.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*(3), 295–355.

Huang, Y. (2011). Types of inference: Entailment, presupposition, and implicature. In W. Bublitz & N. R. Norrick (Eds.), *Foundation of Pragmatics* (pp. 397–424). De Gruyter Mouton.

Johannesson, M. (2000). Modelling asymmetric similarity with prominence. *British Journal of Mathematical and Statistical Psychology, 53*(1), 121–139.

Johannesson, M. (2002). *Geometric models of similarity*. Lund University Cognitive Studies, 90.

Jönsson, M. L., & Hampton, J. A. (2012). The modifier effect in within-category induction: Default inheritance in complex noun phrases. *Language and Cognitive Processes, 27*(1), 90–116.

Keil, F. C., & Batterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning and Verbal Behavior, 23*(2), 221–236.

Kellen, D., & Klauer, K. C. (2020). Theories of the Wason selection task: A critical assessment of boundaries and benchmarks. *Computational Brain & Behavior, 3*(3), 341–353.

Keßler, C., Raubal, M., & Janowicz, K. (2007, November). The effect of context on semantic similarity measurement. In OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". Springer. pp. 1274–1284

Krifka, M., Pelletier, F., Carlson, G., ter Meulen, A., Chierchia, G., & Link, G. (1995). Genericity: An Introduction. In G. Carlson & F. Pelletier (Eds.), *The Generic Book* (pp. 1–124). University of Chicago Press.

Lakoff, G. (2008). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press.

Lawler, J. M. (1973). *Studies in English Generics.* Doctoral dissertation, University of Michigan.

Leslie, S. J. (2008). Generics: Cognition and acquisition. *Philosophical Review, 117*(1), 1–47.

Leslie, S. J. (2012). Generics articulate default generalizations. *Recherches Linguistiques De Vincennes, 41*, 25–44.

Leslie, S. J., & Lerner, A. (2016). Generic generalizations. Stanford Encyclopedia of Philosophy, Stanford, CA.

Leslie, S. J., Khemlani, S., & Glucksberg, S. (2011). Do all ducks lay eggs? The generic overgeneralization effect. *Journal of Memory and Language, 65*(1), 15–31.

Lewis, M., & Lawry, J. (2016). Hierarchical conceptual spaces for concept combination. *Artificial Intelligence, 237*, 204–227.

Lieto, A. (2021). *Cognitive design for artificial minds*. Routledge.

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.

López, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The development of category-based induction. *Child Development, 63*(5), 1070–1090.

MacFarlane, J. G. (2000). *What does it mean to say that logic is formal?* PhD Thesis. University of Pittsburgh

Maddox, T. (1992). Perceptual and decisional separability. In G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147–180). Lawrence Erlbaum Associates Inc.

Mattos, O., & Hinzen, W. (2015). The linguistic roots of natural pedagogy. *Frontiers in Psychology, 6*, 1424.

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). Elsevier.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*(1), 89–115.
Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39–57.
Oaksford, M., & Chater, N. (2009). The uncertain reasoner: Bayes, logic, and rationality. *Behavioral and Brain Sciences, 32*(1), 105–120.
Oaksford, M., & Chater, N. (2020). New paradigms in the psychology of reasoning. *Annual Review of Psychology, 71*, 305–330.
Okabe, A., Boots, B., & Sugihara, K. (1992). *Spatial tessellations: concepts and applications of voronoi diagrams* (Vol. 501). John Wiley & Sons.
Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*(2), 185–200.
Osta-Vélez, M., & Gärdenfors, P. (2020). Category-based induction in conceptual spaces. *Journal of Mathematical Psychology, 96*, 102357.
Osta-Vélez, M., & Gärdenfors, P. (2022a). Nonmonotonic reasoning, expectations orderings, and conceptual spaces. *Journal of Logic, Language and Information, 31*, 77–97.
Osta-Vélez, M., & Gärdenfors, P. (2022b). Analogy as a search procedure: a dimensional view. *Journal of Experimental & Theoretical Artificial Intelligence*. https://doi.org/10.1080/0952813X.2022.2125081
Pelletier, F. J., & Asher, N. (1997). Generics and defaults. In J. van Benthem & A. ter Meulen (Eds.). *Handbook of Logic and Language* (pp. 1125–1177). North Holland.
Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. Penguin Books.
Pollard, P., & Evans, J. S. B. (1987). Content and context effects in reasoning. *The American journal of psychology, 100*, 41–60.
Prasada, S., Khemlani, S., Leslie, S. J., & Glucksberg, S. (2013). Conceptual distinctions amongst generics. *Cognition, 126*(3), 405–422.
Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 130*(3), 323–360.
Rein, J. R., Love, B. C., & Markman, A. B. (2007). Feature relations and feature salience in natural categories. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 29(29), 592–598.
Richie, R., White, B., Bhatia, S., & Hout, M. C. (2020). The spatial arrangement method of measuring similarity can capture high-dimensional semantic structures. *Behavior Research Methods, 52*, 1906–1928.
Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 665–681.
Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Mit Press.
Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*(3), 192–233.
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.
Rumelhart, D. E., & Abrahamson, A. A. (1973). A model for analogical reasoning. *Cognitive Psychology, 5*(1), 1–28.
Ryle, G. (2009). *The concept of mind*. Routledge.
Sadler, D. D., & Shoben, E. J. (1993). Context effects on semantic doamins as seen in analogical solutions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 129–147.
Shastri, L. (1989). Default reasoning in semantic networks: A formalization of recognition and inheritance. *Artificial Intelligence, 39*(3), 283–355.
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.
Sloman, S. A., & Lagnado, D. (2005). The problem of induction. In R. Morrison & K. Holyoak (Eds.), *Cambridge Handbook of thinking and reasoning* (pp. 95–116). Cambridge University Press.
Sloman, S. A., Love, B. C., & Ahn, W. K. (1998). Feature centrality and conceptual coherence. *Cognitive Science, 22*(2), 189–228.

Smith, E. E., Langston, C., & Nisbett, R. E. (1992). The case for rules in reasoning. *Cognitive Science, 16*(1), 1–40.
Sterken, R. (2015a). Generics, content and cognitive bias. *Analytic Philosophy, 56*(1), 75–93.
Sterken, R. (2015b). Generics in context. *Philosopher's. Imprint, 15*(21), 1–30.
Strößner, C. (2022). Criteria for naturalness in conceptual spaces. *Synthese, 200*(2), 78.
Tessler, M. H., & Goodman, N. D. (2019). The language of generalization. *Psychological Review, 126*(3), 395–436.
Thagard, P., & Nisbett, R. (1982). Variability and confirmation. *Philosophical Studies, 42*, 379–394.
Thagard, P., Holyoak, K. J., Nelson, G., & Gochfeld, D. (1990). Analog retrieval by constraint satisfaction. *Artificial Intelligence, 46*(3), 259–310.
Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases* (pp. 84–98). Cambridge University Press.
Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327–352.
Ungerer, F., & Schmid, H. J. (2006). *An introduction to Cognitive Linguistics*. Routledge.
Van Rooij, R., & Schultz, K. (2019). Generic sentences: representativenesss or causality?. *Proceedings of Sinn Und Bedeutung*, 23(2), 409–426.
Verheyen, S., & Égré, P. (2018). Typicality and graded membership in dimensional adjectives. *Cognitive Science, 42*(7), 2250–2286.
Wheeler, D., Tripp, E. E., & Natarajan, B. (2022). Semantic communication with conceptual spaces. *IEEE Communications Letters, 27*(2), 532–535.