Conceptual Coherence

Matías Osta-Vélez[*] and Peter Gärdenfors[**]

[*] Department of Philosophy, Heinrich Heine University Düsseldorf

Institute of Philosophy, FHCE, University of the Republic

[**] Department of Philosophy and Cognitive Science, Lund University

October 2024

Abstract

We propose an explication of conceptual coherence in terms of the covariational structure of concepts or how clusters of properties systematically co-occur across category exemplars. Using the theory of conceptual spaces combined with ideas from Principal Component Analysis, we show that a concept's perceived coherence relates to how easily its attribute structure can be reduced to simpler representations. Our approach contrasts with previous accounts that ground coherence in similarity or intuitive theories. We discuss the relationship between coherence, uncertainty, and induction and apply our framework to the conjunction fallacy.

1. Introduction

One fundamental aspect of concepts is that they are *general*. That is, they group *different representations under one common representation* (Kant 1997, B93 = A68). The generality of concepts promotes cognitive efficiency by allowing non-identical stimuli to be treated as equivalent for specific purposes. However, the generality of a concept must be weighed against its informational function. The concept must capture shared information among its instances while remaining broad enough to cover a range of them.

Our conceptual system includes a variety of concepts that exhibit different levels of generality. For example, broad concepts such as LIVING BEING encompass a wide array of instances but provide relatively little specific information. In contrast, narrower concepts like PUFFIN include fewer instances but offer more precise details. This leads to two questions: How can we account for the trade-off between generality and specificity within a concept? How does the diversity of a concept's exemplars influence its application?

These problems are seldom addressed in theories of concepts in psychology, with the notable exception of Rosch's work. They underpin a notion that is becoming increasingly important in the field and is the central concern of this paper: conceptual (or category) coherence. Conceptual coherence refers to the extent to which properties associated with a particular concept are perceived as *belonging together* or, in other words, to the *perceived unity* among instances of a concept based on shared features, functions, or other underlying characteristics (Smith & Medin 1981). From this perspective, coherence can be considered a "global" or "configural" property of concepts, which concerns the concept when taken as a whole and consequently requires some form of 'holistic' processing.[1]

The coherence of concepts is often reflected in the confidence they are used across cognitive tasks, especially in inductive tasks (e.g., Lassaline and Murphy 1996). A key example is the "preferred level of induction" phenomenon (Sloman and Lagnado 2005, p. 106), which

[1] Here, we follow a common distinction from gestalt psychology between *local* and *holistic* processing, two ways of understanding perceptual phenomena. Local processing involves focusing on individual elements of a visual scene, such as specific shapes, colors, or objects, often without regard to the overall structure. In contrast, holistic processing interprets the scene as a whole, recognizing patterns or configurations that emerge from the interaction of its elements (see Wagemans et al. 2012).

refers to the tendency to favor basic-level categories (e.g., CHAIR or DOG) over more general ones (e.g., FURNITURE or MAMMAL) during categorization and inference.

There is strong evidence that natural concepts (e.g., CAT or GRASS) are perceived as more coherent than artificial ones (e.g., BIKE or GROCERIES) and even more so than ad-hoc concepts (e.g., THINGS TO BRING CAMPING or THINGS TO SAVE FROM A FIRE). Barsalou (1983) suggested that because ad hoc concepts are highly context-dependent and less frequently encountered than natural ones, they are not only used with more resistance in inductive tasks, but they are also more difficult to retain in memory.

The primary focus of this paper is to explore what is computed when estimating the coherence of a concept, an aspect we believe has not previously been clearly articulated in the cognitive psychology literature.[2]

1.1 Coherence and Similarity

In addressing the question of what underlies conceptual coherence, cognitive psychologists have posited two different responses. The first asserts that conceptual coherence is grounded in semantic similarity. Simply put, we perceive a concept as coherent when its instances share sufficient similarities, forming a 'natural grouping'. Rosch and Mervis (1975) provided evidence that members of basic-level concepts tend to have many features in common, surpassing those of their corresponding superordinate concepts and closely mirroring those of

[2] In this article, we focus solely on the problem of coherence within what we call "object concepts." These are concepts that represent collections of entities denoted by a noun or phrasal nominal (e.g., "dog," "old lawyer," "things to take to the beach"). They differ from properties or attributes, typically linked to adjectives in natural languages, and from abstract concepts (e.g., "freedom," "dignity," "inflation") that cannot be easily analyzed in terms of dimensions. As a notational criterion for the remainder of this article, we will use small caps for concepts and italics for their properties or attributes.

their subordinate concepts. They concluded that basic-level concepts optimize intra-category similarity while concurrently minimizing inter-category similarity, distinguishing them from others in the taxonomic hierarchy. Consequently, they posited that the inherent coherence of basic-level concepts underpins their preferential status in induction and categorization tasks.

However, the similarity-based view of coherence faces several challenges. The most pressing of these is establishing a criterion for identifying which attributes should be considered relevant in similarity judgments. Prima facie, shared attributes can be identified between any pair of objects (e.g., a zebra and a crosswalk), and this issue becomes even more complex when including negative predicates as shared features. For instance, when Lewis Carroll asked why a raven is like a writing desk, part of his whimsical answer was that "it is never put with the wrong end in front." While this negative property is common to many objects, it is unhelpful for categorization and reasoning. Furthermore, even if we find a method for determining which attributes are relevant to a concept, we still need a way to assign the relative importance (or 'weights') of each attribute for similarity comparisons—a task that has proven challenging (see Gelman and Williams 1988). Given these difficulties, several psychologists have suggested that similarity may be a by-product of conceptual coherence rather than its foundational principle (e.g., Medin and Wattenmaker 1987, p. 28).

1.2 The Theory-theory

The second approach to the coherence problem addresses the limitations of similarity by positing that concepts are embedded in *intuitive theories* (e.g., Gopnik and Meltzoff 1997, Murphy and Medin 1985). An intuitive theory is defined as "a system of interrelated concepts that generate explanations and predictions in a specific domain of experience" (Slaughter and Gopnik 1996, p. 2967). Such theories determine the parameters, like relevance and weight of attributes, that constrain semantic similarity and guide concept formation. Moreover, intuitive

theories provide concepts with causal structure by building causal links between the features. For example, attributes such as *gills*, *fins*, and *streamlined bodies* may be particularly important for the concept FISH because they are causally related to one another in a person's naive theory of how fish swim. Endowing categories with a causal structure enhances their coherence by explaining why certain attribute groups are perceived as "holding together."

This approach, commonly known as "Theory-Theory", is not without its challenges. One major issue is that the most basic definition of a theory is a logically structured set of propositions encompassing both theoretical and observational language (Giere, 2000) — definitions in the psychological literature tend to be more intricate (see Gopnik and Meltzoff 1997, pp. 32-41). However, since propositions are composed of concepts, the definition of propositions inherently relies on a prior understanding of concepts. As a result, defining 'concept' in terms of 'theory' appears to reverse the logical explanatory order, creating a risk of circularity.

Moreover, even if we concede that concepts are rooted in intuitive theories, the genesis of these theories remains a mystery (Sloutsky 2003). Finally, akin to the possibility of having multiple criteria for assessing similarity between two objects, different theories can be consistent with the same dataset (see Turnbull 2018). Scientists rely on meta-theoretical criteria to choose the most suitable theories from available alternatives. However, the reason we have a particular intuitive biology for our biological categories rather than another framework highlighting distinct causal structures remains unclear.

1.3 Outline

In this paper, we will defend an alternative approach that eschews the notion of theory to explain coherence. Our central argument is that a concept's perceived coherence hinges on identifying attribute clusters that exhibit systematic covariation across its diverse exemplars.

For example, in the concept BIRD, the consistent relationship between *beak shape* and *feeding habits* across various species illustrates this kind of systematic covariation. The greater the covariance observed within a concept's feature space, the stronger its perceived coherence. We will then demonstrate how these ideas can be implemented and systematized within the framework of the theory of conceptual spaces. Then, we will introduce a method to measure the coherence of a concept using ideas from Principal Component Analysis (PCA).

The remainder of this article is as follows: The subsequent section offers a concise review of prior attempts to explicate the notion of conceptual coherence. Section 3 introduces the theory of conceptual spaces, laying the groundwork for our formal approach to coherence. In Section 4, we delve into the interplay between the heterogeneity of a concept and its coherence, and we introduce a method based on principal components to measure the latter within the realm of concept spaces. Section 5 explores the relation between coherence and induction and uses the developed ideas to explain the conjunction fallacy.

2. Previous attempts to model coherence

While the primary aim of this article is to provide a precise definition of a notion frequently employed in an unclear manner in psychology, we ultimately view the problem of coherence as another facet of the general puzzle of induction (see Kornblith 1995). In this regard, our perspective aligns with Nelson Goodman's ideas on how the 'quality' of a predicate influences its inductive power. Goodman (1983) was likely the first to observe that induction isn't merely a formal mechanism but hinges on the content of the predicate being projected and its position within a broader system of predicates. Rather than discussing the coherence of a concept, Goodman talked about 'well' and 'ill-behaved' predicates. A well-behaved predicate is easily

projectable and deeply *entrenched* within the overarching system of predicates. Clear examples include animal categories.

Conversely, an ill-behaved predicate is weakly entrenched, possessing few or no 'parent predicates', directly impacting its inductive power or projectability. Clear examples of this are ad-hoc concepts. Items falling under this concept might be vastly dissimilar, correlations between features are limited, and it lacks clear superordinate concepts, i.e., it doesn't fit neatly into a conceptual taxonomy. The result is that the inductive power of such concepts is markedly low.

Goodman's distinction aligns with our characterization of high and low-coherence concepts. However, although he introduced insightful notions, such as 'entrenchment,' he did not fully elaborate on these ideas, nor did he provide a formal articulation of his concepts. To our knowledge, the first — and only — formal model of conceptual coherence was proposed by Paul Thagard (2002). This model is a specific application of his broader framework, designed to accommodate various types of inputs, not just concepts, including beliefs, hypotheses, percepts, and propositions.[3]

The central premise of Thagard's model is that evaluating coherence can be seen as a constraint satisfaction problem. We start with a set of elements E that can be related to each other based on coherence or incoherence relationships. Coherence relations between elements (e_i, e_j) in E are mapped into a set of positive constraints, while incoherence relations are mapped into a set of negative constraintsC^-. The elements in sets C^+ and C^- have assigned

[3] Numerous probabilistic models of coherence have been developed within the realm of formal epistemology (see, for instance, Douven and Meijs 2007 or Hartmann and Trpin 2023). However, these models address a phenomenon distinct from our primary interest here. Specifically, they aim to measure the coherence between sets of propositions, rather than the coherence of the internal structure of categories.

weights w_{ij} corresponding to their relative importance for the problem. The objective is to identify a subset of elements from E that satisfy the maximum number of constraints, that is, to find a subset that *makes more sense* than any other subset of E. We then partition E into a set A of accepted elements and a set R of rejected elements, and we compute the weight of the partition $w(A, R)$ as the sums of the weights of the satisfied constraints, where a constraint is satisfied if for any element(e_i, e_j) in C^+or C^- , $e_i \in A$ iff $e_j \in A$ or , $e_i \in R$ iff $e_j \in R$. The coherence problem is resolved when the partition with the maximum weight is identified.

Thagard's model has been implemented in connectionist networks and has found various applications (e.g., Thagard et al. 2002). However, it falls short of explaining the nature of conceptual coherence. The reason is straightforward: explaining conceptual coherence requires clarifying the nature of the relationships between a concept's components that determine its overall coherence. In Thagard's model, this would be equivalent to describing the constraints in sets C^+or C^-, an issue that Thagard intentionally avoids. Additionally, Thagard's model is unclear on which entities can engage in coherence relations. The elements of *E* can represent a single concept and its internal features (as in the model we propose) or a set of thematically related concepts like *dog-leash-veterinarian*. The type of conceptual coherence we analyze, which is the one commonly used in cognitive psychology, is restricted to cases of the former type.

3. Conceptual spaces as a framework

Conceptual spaces (Gärdenfors 2000, 2014) have been developed as a research program in cognitive semantics, studying the structure of concepts and their interrelations through geometrical methods. This approach builds on two key ideas regarding the composition and structure of concepts and properties: (i) they are composed of clusters of quality dimensions,

many of which are generated by sensory inputs such as color, size, and temperature, and (ii) they possess a geometric or topological structure, resulting from the integration of the specific structures of these dimensions.

Quality dimensions can be either *integral* or *separable*. Dimensions are integral when assigning a value to an object on one dimension necessarily entails assigning a value on another dimension (Maddox, 1992). For example, it is not possible to attribute a value to the pitch of a tone without also attributing one to its loudness.

A *domain* is defined as a set of integral dimensions that are separable from all other dimensions. For instance, color properties are composed of three fundamental parameters of color perception: hue, saturation, and brightness (Gärdenfors 2000, 2014). Any perceived color can be mapped to specific values along these dimensions. More generally, different colors can be described as *regions* of possible values across these three parameters.

A central idea of this theory is that natural properties (like colors) correspond to convex regions of a single domain (Gärdenfors 2000, p. 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.

A conceptual space consists of a collection of one or more domains, equipped with a distance function (or *metric*). The choice of distance function can vary; the most common is the Euclidean metric, though Manhattan and polar metrics may also be appropriate in different contexts (see Shepard 1964; Johannesson 2002; Gärdenfors 2014).

Similarity among concepts and objects is defined as a monotonically decreasing function of their distance within the space (Shepard 1987). This contrasts with Tversky's (1977) approach, which compares the number of properties two objects share with the number of properties where they differ.

Many predicates in natural language, particularly those expressed by nouns, cannot be defined within a single domain but instead as clusters of properties. This distinction leads us to

categorize predicates as either *properties* or *concepts*. Properties are convex regions within single domains, whereas concepts are convex regions spanning a set of interconnected domains (Gärdenfors 2000, Sec. 4.2.1). For most concepts, the domains that compose them are correlated in various ways. For instance, in the case of the concept FRUIT, properties such as *size* and *weight*, or *ripeness*, *color*, and *taste*, tend to covary. These covariations generate expectations crucial for inferential processes that build on the structure of semantic representation.

As an example, consider a simplified conceptual space for FRUIT, defined by five dimensions representing key properties of fruits: *color*, *taste*, *ripeness*, *texture*, *size*, and *shape*. The "fruit space" is the Cartesian product of these six dimensions. The concept APPLE occupies specific subregions within this space, corresponding to the range of possible properties for instances of apples, as well as correlations between these dimensions (see Figure 2).

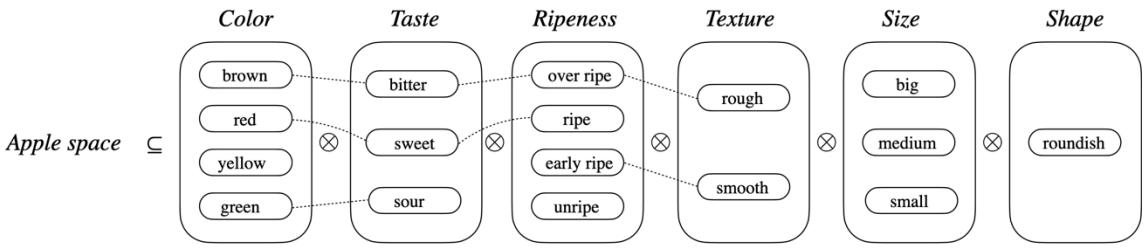

FIGURE 1. The concept of an apple as a subregion of 'fruit space'. The dotted lines represent correlations between properties of the concept APPLE

An important advantage of representing concepts this way is that it naturally accounts for the prototypical structure of categories (Rosch 1975, 1983; Gärdenfors 2000, Lakoff 2008). When concepts are defined as convex regions within *n*-dimensional spaces, a specific point in each region can be interpreted as the prototype for the corresponding property or concept.

Conversely, given a set of prototypes $p_1, p_2, \ldots, p_n$ and a Euclidean metric, a set of n concepts can be delineated 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 corresponds to the Voronoi tessellation, an example of which is illustrated in Figure 2. Thus, assuming a metric is defined on the subspace under categorization, a set of prototypes will generate a unique partitioning of the subspace into convex regions by this method.

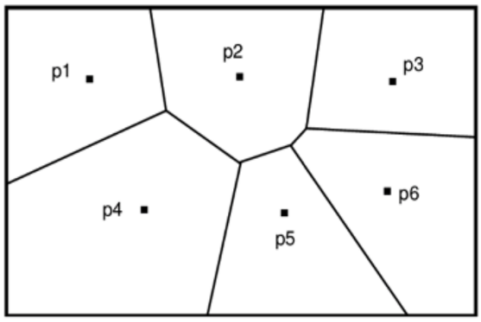

FIGURE 2. Voronoi partitioning in a 2-dimensional space.

Within this framework, objects are viewed as instances of concepts and are mapped to points in the space, while concepts are represented as regions (convex sets of points). This representation accommodates graded membership and degrees of typicality (Rosch et al. 1976; Hampton 2007), meaning that objects can be represented as more or less typical instances of categories depending on their position relative to the prototype.

4. Measuring coherence

4.1 Heterogeneity and Uncertainty

Before introducing our measure of coherence, it's important first to discuss a closely related concept: concept (or category) heterogeneity. Psychologists often describe this as the balance

between variability and similarity among exemplars within a concept or category (see Gelman 2003; Brandone 2017). Concept heterogeneity refers to the volume of the feature space spanned by a concept, indicating the degree of diversity among its members. For instance, PIPE WRENCH is a highly homogeneous concept, with exemplars showing minimal variation in key dimensions like *function*, *shape*, and *material*. As a result, these exemplars are highly similar, making the concept compact. In contrast, a concept like HANDBAG is more heterogeneous, including objects that vary in shape, size, material, and function, yet are all recognized as handbags.

The heterogeneity of a concept is influenced by its position within a concept hierarchy. A superordinate concept like MAMMAL is more heterogeneous than any of its subordinates (e.g., DOG, TIGER, COW) because it encompasses them all along with their diverse properties. This has significant implications for the inferential use of concepts, particularly in contexts involving prediction and inductive reasoning. Specifically, the greater the variability within a concept, the weaker its predictive power. For example, from the statement "x is a mammal," we can infer that *x* likely has fur or hair, is warm-blooded, and produces milk. However, we cannot predict its size (which could range from a tiny shrew to a massive blue whale), diet (herbivore, carnivore, or omnivore), or habitat (land, water, or air). In contrast, from "*x* is a dog," we can make more specific predictions: *x* has four legs, a tail, barks, is primarily carnivorous, is domesticated, and likely lives close to humans. This illustrates how the more specific concepts allow for stronger, more detailed predictions than the broader, more heterogeneous ones (for a more detailed explanation of this phenomenon, see Thagard and Nisbett 1982, Sloman and Lagnado 2005, Brandone 2017).

To better understand the relationship between heterogeneity and inference, consider the link between categorization and uncertainty. Categorization acts as a mechanism that maps an input to a categorical output (a concept). However, as the heterogeneity of the matched concept

increases, this hypothesis space expands, influencing both the quality and quantity of predictions we can make about the input.

This idea fits naturally within the framework of conceptual spaces. The volume of a concept in a conceptual space —defined as the size of the region representing the concept— is directly proportional to the variability of properties and the number of relevant dimensions it contains. When we categorize an input, we essentially represent it as a potential point within this conceptual space. Hence, the volume (and, by extension, the heterogeneity) of the concept influences the number of possible points the input could occupy within that space. Gathering more information about the input, whether by pinpointing specific properties or by recategorizing it into a more specific concept, narrows down the volume of potential points, thereby reducing our uncertainty about the input.

How do heterogeneity and coherence interact within a concept? In our perspective, the coherence of a concept pertains to the level of interconnectedness or covariation among its properties. We use the term "covariational structure" to denote the aggregate of all covarying dimensions known to an agent familiar with the concept. Conversely, heterogeneity reflects the diversity and range of properties within that concept. This diversity not only characterizes the concept but also defines the potential range of its coherence —it establishes the boundaries for coherence. Essentially, the greater the diversity of a concept, the broader the potential scope for its coherence. This spectrum ranges from a loosely connected concept to one of maximal coherence, where knowledge of one property allows prediction of the rest.

4.3 Measuring Coherence

In coherent concepts, clusters of properties' hang together,' indicating that they covary. Within the semantic domain of animals, for instance, a creature's size might covary with specific attributes like *lifespan*, *diet*, or *habitat*. Noticing that larger mammals often have longer

lifespans than smaller insects is a simplistic illustration of this covariation. Similarly, a bird's beak shape can often hint at multiple correlated attributes like its *diet*, *habitat*, *nesting behavior*, and even *mating calls*. For instance, observing a bird with a long, slender beak might suggest that it primarily feeds on nectar. This could further correlate with living in colorful, flower-rich habitats, crafting intricate hanging nests, and producing melodious calls to communicate.

Research by Younger (1990), Billman and Knutson (1996), and Hayes et al. (1996) shows that humans excel at detecting covariations across various domains. This ability likely stems from our evolutionary history, where recognizing patterns in nature was crucial for survival, enabling better predictions of hazards, food sources, and suitable habitats. Through natural selection, humans have developed an innate ability to recognize such grouped relationships. This evolutionary context helps explain the basic-level concepts described as distinctive clusters of covarying properties in Rosch's (1975) prototype theory (Holland et al. 1986, pp. 183–4).

From this perspective, measuring coherence depends on assessing the covariational structure of a concept, which is not straightforward. Our proposal is to address this challenge using ideas from Principal Component Analysis (PCA). Specifically, we assert that the coherence of a concept depends on how much the first principal component of the set of objects under that concept reduces overall variation within the set. To elaborate on this proposal, we must first explain what principal components are.

If a concept is represented in a conceptual space, its instances form a set of points (a "dataset") within that space. This set can be more or less 'organized.' To analyze this organization, we can use PCA, which helps identify the underlying structure of the dataset. Specifically, PCA identifies principal components—directions in the space along which the data varies the most.

The first principal component is the line in the space that, when the data points are projected onto it, maximizes the reduction in overall variance. In other words, this component represents the dimension that provides the most comprehensive 'explanation' of the data (see Jolliffe 2002). The second principal component, orthogonal to the first, accounts for the remaining variation, and further components follow in a similar manner. However, each additional component typically explains progressively less variance unless the dataset is nearly random.

The relevance of this to conceptual coherence is that the length of a principal component reflects how much variance in the dataset it explains. A longer first principal component indicates a more organized or coherent concept because more of the variance is captured by that component. In contrast, if the variance is spread across many components, the concept is less structured or coherent.

As an illustrative example, assume that the concept STRAWBERRY can be represented by merely three dimensions: *color*, *taste,* and *size*. Suppose a number of observations of strawberries result in the dataset depicted in Figure 3. The first principal component of the set is the long vector that extends diagonally upwards to the right from the mean of the dataset. Orthogonal to that vector, we find a shorter vector representing the second principal component and an even shorter vector representing the third principal component. In this example, the first principal component captures considerably more variance than the other two combined and can be used to re-express the data in an efficient manner — without significant loss of information. The dimension identified by the first principal component can be interpreted as a representation of the 'ripeness' of the strawberries.

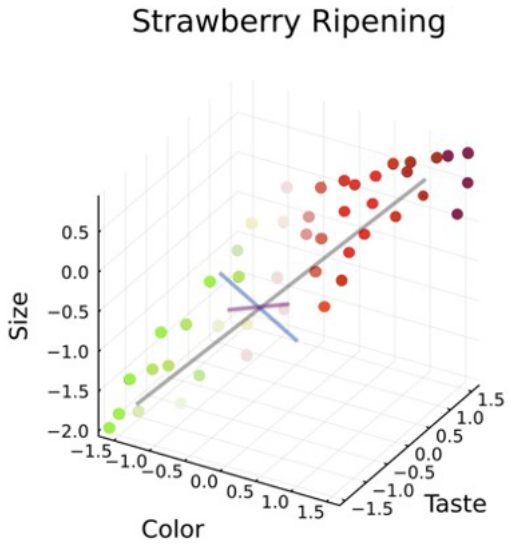

FIGURE 3. An illustration of the first principal components of a data set.

Note that the very fact that we can consistently talk about the 'ripeness' of fruits suggests that we identify underlying multivariate structures. These structures reveal the presence of interrelated properties that, when analyzed collectively, enable dimensionality reduction by encapsulating the variance of individual properties into a singular, cohesive descriptor. Essentially, the use of multifaceted descriptors in language, such as 'ripeness,' serves as a mechanism for achieving both communicative and cognitive efficiency. By compressing diverse information into a single term, language users can convey and understand intricate ideas without becoming bogged down in details. This aligns with the principle of information economy, which aims to transmit the most information with the least effort.

Our main hypothesis is that the principal component (or the first components) of a concept can be used to characterize coherence. We propose using the proportion of the variance explained by the first principal component as our measure. When all points in the dataset are positioned on a straight line, that line will represent the principal component, yielding a maximal coherence value of 1. Conversely, in a scenario where the data points are scattered

randomly, any line might be considered the principal component. However, such a line will not significantly reduce the variance, resulting in a coherence value of 0.

In the example above, only three dimensions/domains were involved. For most natural concepts, the number of domains is very large. These domains are more or less salient for the concept, with many being only marginally relevant or not relevant at all. For example, even though all dogs have a temperature, the variation of this variable is typically so small that this domain has no influence on the principal component of the concept of dogs. In order to make our central hypothesis more precise, we need to get into a little bit of technicalities. The standard method for principal component analysis proceeds as follows (Jolliffe, 2002):

(i) Standardize the range of continuous initial variables. This step ensures that all variables are on the same scale and that no variable dominates the analysis due to its scale.[4]

(ii) Compute the covariance matrix to identify correlations. This matrix provides a measure of how much each pair of variables varies together.

(iii) Compute the eigenvectors and eigenvalues of the covariance matrix to identify the principal components. The eigenvectors represent the directions of maximum variance in the data, and the eigenvalues represent the magnitude of these variances. The principal components are the eigenvectors of the covariance matrix, ordered by their corresponding eigenvalues.

Now, the first step would imply that all dimensions relevant to a concept are given equal weight. This essentially involves calculating correlations rather than covariations. In the conceptual space representation of a concept, dimensions exhibit varying salience, and these salience values should be utilized when calculating the principal component of the associated

[4] In its standard form, PCA is applied to data described in continuous variables. However, most of our concepts combine continuous variables with categorical variables. There are variations of PCA, like Multivariate Analysis (Chavent et al. 2014), that apply the same principles to mixed data sets.

data set. This also indicates that the ranges of the variables are not standardized. For instance, in the earlier example, the range of the color dimension is slightly larger than that of the sweetness dimension, yet it may be more intuitive to assign a higher salience value to sweetness than to color when comparing the dimensions. Consequently, our proposed method for determining principal components focuses on calculating covariations using specified salience weights rather than relying on correlations based on standardized ranges of the data set dimensions.

Considering that objects in conceptual spaces are represented as points, these points can be interpreted as vectors extending from a defined 'origin'. We refer to this origin as the 'universal prototype'—the prototype of THING. This prototype may be elusive, so when calculating covariance, the origin could be considered the mean of the positions of all the objects against which variance is measured.

The role of the origin as a way of calculating the covariance is evident in concepts that exhibit little or no variation in most of their dimensions. Many animal concepts, for example TIGER, would be of this kind. For such categories, there will be little or no correlations between the dimensions. However, because covariance is assessed in relation to the entire conceptual space, it remains strong. This is the reason why we use covariance rather than correlation to determine the coherence of a concept.

There are several aspects of this coherence measure that should be noted. First, the measure is dependent on the salience values of the domains. If a domain has low salience, then all volume related to it will be small, and, consequently, its contribution to the total coherence of the concept will be small. For example, when WHALE was reclassified from FISH to MAMMAL, more salience was given to biological domains and less to ecological domains (Gärdenfors 2000, Section 6.4). As a consequence, WHALE became a more coherent concept.

Second, the measure is independent of any probabilities of the properties — only the volume of the intersection of the regions matters. We see that as an advantage of the conceptual spaces framework, as it allows us to assess coherence without needing to consider probabilities.

Third, for a concept with several salient domains, the coherence value will, in general be higher than for a concept with a few salient domains. This explains why concepts for natural categories such as BIRD will have a high coherence value since it has several salient domains that covary, while concepts for artifacts such as CHAIR or CLOCK will have low coherence values. For artifacts, physical properties are often non-salient and it is only properties related to function that matter.

Fourth, if a concept has clusters of covarying properties, for example *flying*, *feathers*, *wings*, *beak* for the concept BIRD, then the coherence value will be high. Again, such clusters are typically not found for artifacts. Clusters of correlated properties will also be helpful in learning the meaning of a concept, which may be an explanation of why coherent concepts are easier to learn than non-coherent ones (Billman and Knutson 1996, Kornblith 1993, Shipley 1993).

5 Coherence, Causality, and Induction

5.1 Coherence and Causality

Some authors challenge the idea that coherence is solely grounded in covariational learning, arguing instead that causal knowledge plays a primary role. In a series of experiments, Malt and Smith (1984) and Ahn et al. (2002) showed that typicality judgments about exemplars of a concept were stronger for those that included pairs of casually correlated properties. In their experiments, they provided descriptive exemplars of the concept BIRD. Besides sharing general properties like *has wings*, some exemplars had the properties *lives near the ocean* and *eats fish*,

while others *is white* and *eats fish*. Although both property pairs exhibited a similar degree of covariation, subjects judged as more typical the exemplars that had the pair lives near the ocean and eats fish because they saw these properties as causally related. They concluded that semantic judgments depend not only on pairwise covariations but also on causality judgments, which is consistent with theory-theory and has implications for the notion of coherence.

There is, however, an alternative explanation (see also Rogers and McClelland 2004). Participants might give more weight to properties they perceive as causally linked because they coherently covary with multiple other properties. For instance, *lives near the ocean* and *eats fish* apply to many shorebirds sharing other common traits like *dives* and *can swim*. Conversely, *is white* and *eats fish* could apply to various birds, for example, herons, making this combination less coherent. The implication is that causality might not be essential to explain coherence. Instead, the coherence of two properties may stem from their covariation with other properties, contributing to the principal components and thereby enhancing coherence. This aligns with the measure we propose.

Moreover, our perception of the connections between property pairs within a concept may be influenced not only by their covariation but also by the relationships among properties across similar concepts within the same domain. For instance, individuals might perceive the pair *is white* and *eats fish* as less correlated because they recognize that color and feeding habits are not consistently related across different animal concepts. In contrast, the pair *lives near the ocean* and *eats fish* may be viewed as more correlated, as habitat and feeding habits often exhibit a strong relationship across various animal concepts. This broader perspective suggests that concept coherence is shaped by the statistical structure of the environment, not just within a single concept but across related concepts.

In a similar line of argumentation that takes causality as the central factor behind coherence, Rehder (2017) and Rehder and Kim (2006, 2010) explored and discussed a

'coherence effect' in people's judgments of the coherence of instances of concepts: people are more likely to judge an exemplar of a concept as good if it does not violate the patterns of causal relationships between its features that were previously learned —the 'causal laws' of the concept (Rehder and Hastie 2001).

As an example, consider the concept TROPICAL FROG: Suppose people have learned that being poisonous (feature A) causes brightly colored skin (feature B) in these frogs. Given this causal relationship, people judge exemplars of frogs that are not poisonous and not brightly colored (¬A¬B) as better category members than exemplars that are either not poisonous but brightly colored or poisonous but not brightly colored (¬AB or A¬B). This occurs because the ¬A¬B exemplars preserve the learned causal structure (no poison, therefore no bright color). In contrast, exemplars ¬AB and A¬B violate this structure, as their features don't align with the expected causal relationship.

Our approach can explain this effect if we look at the role of loadings and scores in the PCA analysis of the concepts. In PCA, the loadings are the elements of the eigenvectors obtained from the covariance matrix of the original variables. Each variable has its own loading on some principal component. This loading indicates how much that variable contributes to the component and how well the component explains the variability in that variable. The scores, on the other hand, are the transformed value of a data point along a principal component. In other words, it represents the projection of an original data point onto the principal component space.

When people learn that feature A causes feature B in a concept, it reinforces their expectation that these features will co-occur, either both present (AB) or both absent (¬A¬B). This learned causal knowledge shapes the correlational structure of the concept, with A and B becoming strongly correlated across exemplars. In a PCA of the concept, A and B would load heavily on the same principal component, likely the first component if they are central or typical

features. Exemplars (i.e., data points) with features AB will score high in the principal component where A and B have high loadings.

The high absolute value of the score indicates that the exemplar strongly aligns with the causal-correlational structure captured by the component. Conversely, an exemplar that lacks both features A and B (¬A¬B) will also score high on the same component but with the opposite sign compared to the AB exemplar. This is because the absence of both causally related features is also consistent with the causal-correlational structure of the category. Finally, exemplars that have only one of the features (A¬B or ¬AB) will have scores closer to zero on that component, indicating that they do not align well with the causal correlational structure of the concept. These exemplars will be considered less coherent because they violate the expected co-occurrence of features A and B.

To be clear, we do not claim that causality is irrelevant to conceptual coherence. But we believe that, if we set aside expert knowledge, most of our causal knowledge about inter-feature relations is probably 'shallow'. For example, we know that the curved shape of a boomerang or the size of a bird's wings play a causal role in (or are *enabling conditions* of) the respective flight patterns, but few competent really grasp the underlying physical mechanisms that govern these causal relationships.

We believe that much of our knowledge of the causal structure of concepts has been learned from causal generics (for example, "birds can fly because they have wings") that do not include explanations of underlying causal mechanisms but that still determine our expectations about the patterns of correlations between the properties of a concept (see Gärdenfors and Osta-Vélez 2024). The causal generic "A causes B in concept X" is a shortcut to a type of statistical knowledge that should have been learned by observing a large volume of exemplars of a concept, something that is often difficult.

In summary, we believe that causal knowledge is certainly crucial for the perception of conceptual coherence but that this is a phenomenon limited to expert knowledge, whereas, in our everyday knowledge, that is, for "folk" concepts, the perception of coherence is based on statistical mechanisms that identify sets of candidate causes or enabling conditions for different attributes.

5.2 Coherence and Induction

Empirical studies consistently demonstrate a positive correlation between coherence and the confidence with which concepts are utilized in inductive tasks. In an influential study, Gelman (1988) demonstrated that children exhibit greater sensitivity to conceptual coherence, making more inductive inferences about natural kinds as opposed to artifact concepts. For example, TIGER allows for a wealth of inductive inferences, while CLOCK does not, since different kinds of clocks have few properties in common except for keeping time. The latter often exhibit fewer correlations; thus, their principal component would account for less variance.

Rehder and Hastie (2004) further noted that inductive generalizations are stronger when based on coherent concepts. They proposed that the typicality effect observed in category-based induction can be interpreted as a coherence effect: atypical concept members support weaker generalizations than more typical members because they violate expected correlations associated with that concept. For instance, penguins are considered atypical birds; while they possess feathers and wings, they do not fly, contributing to their weaker inductive generalizations.

Our approach to coherence facilitates a more direct and intuitive explanation of its relationship to induction. Consider, for instance, the phenomenon of within-category induction, where we predict new properties of a categorical input from some already known property (for example, "x is a red apple; thus, x is sweet"). This type of induction clearly relies on our

understanding of correlations. The more coherent our categories are —meaning the richer their correlational structure—the better they enable us to make such inductive inferences, thereby helping us manage the uncertainty inherent in basic conceptual information (cf. Osta-Vélez and Gärdenfors, 2022).

Coherence seems to be important for between-categories induction too. For example, in category-based induction, people rely heavily on the overall perceived similarity between two concepts to project a property of one to the other (Douven et al., 2023). However, it is not always about overall similarities. Sometimes, specific clusters of correlated properties take precedence. Heit and Rubinstein (1994) found that while drawing inferences about behavioral patterns of animals, such as nocturnal feeding habits, subjects found the pairing of tiger and hawk more compelling than chicken and hawk, even if the overall category similarities might suggest otherwise.

6. Coherence, social concepts, and the conjunction fallacy

So far, we have sought to explain why our cognitive system strives to construct and prioritize coherent concepts, using natural kind concepts as the primary example. However, we also claim that social concepts (for example, BANKER, GUITARIST, CONSERVATIVE, FEMINIST) and their use in categorization are especially influenced by coherence and, consequently, by correlations (Patalano et al. 2006). Experimental evidence supports this idea. For instance, Nguyen and Chevalier (2015) discovered that 5-year-olds favor coherent concepts when making inductive decisions about social concepts. In one of their experiment, participants were told the following: "Baseball players like apples, and board game players like bananas. This is Pat. Pat is both a baseball player (coherent concept) and a board game player (incoherent concept)." When asked

whether they believed Pat would prefer an apple or a banana, most chose the fruit linked to the coherent concept.

Social and natural concepts diverge significantly in their basis of categorization. Natural concepts are typically defined by observable features, both behavioral and physical, and remain relatively consistent across various cultural and social contexts. Conversely, social concepts are defined by socially constructed characteristics or roles; they greatly vary regarding the types of features they denote —from patterns of observable behavior to doxastic and ideological dispositions— and they show less stability over time and across cultural contexts compared to natural concepts. More importantly, natural and social concepts have quite different informational profiles. A natural concept provides rather precise information on the distribution of properties of exemplars and prevents cross-classification with any other concept at the same hierarchical level. Social concepts, on the other hand, allow for a great deal of cross-classification. Individuals can belong to multiple social concepts simultaneously, and even if stereotypes exist to navigate the great diversity that we face during social categorization, there is generally no social concept that exhaustively determines the other properties that the categorized person might have.

This leads us to speculate that when we learn and reason about people, we do something that resembles "inference learning"[5] (Jones and Ross 2011, Yamauchi et al. 2002); that is, we try to predict the new concepts into which an individual might be classified, drawing from our general knowledge about correlations and selecting those that best match our prior understanding of the person. In essence, if we have certain prior knowledge about a person *x*,

[5] Inference learning is often contrasted to "classification learning," which consists of predicting a concept label by identifying diagnostic features of exemplars.

we expect *x* to possess properties that align positively with this knowledge and lack properties that are negatively correlated with it.

This idea offers a straightforward explanation of the conjunction fallacy from a 'coherentist' perspective. In the famous experiment by Tversky and Kahneman (1983), participants were presented with a description of Linda, a 31-year-old woman who was single, outspoken, and highly intelligent. She majored in philosophy and, as a student, she was deeply engaged with issues of discrimination and social justice, also participating in anti-nuclear demonstrations. Participants were then asked to assess the likelihood of two scenarios: Linda being a bank teller or Linda being both a bank teller and a feminist. Despite the statistical principle that the probability of two events occurring together is always less than or equal to the probability of either event occurring alone, most participants selected the latter scenario, thereby exhibiting the conjunction fallacy.

From our perspective, the fallacy arises because the property of being a feminist is strongly correlated with the set of attributes previously described about Linda. The scenario of Linda being a feminist bank teller enhances the coherence of our concept of Linda compared to her being merely a bank teller. In other words, if our aim (or the initial presupposition) is to maintain relative coherence in our concepts about individuals, then it seems more reasonable to predict that Linda is a feminist bank teller. Predicting her as only a bank teller increases the uncertainty regarding potential attributes Linda might possess that are positively correlated to being a bank teller (e.g., not being a feminist) but negatively correlated to our prior knowledge about Linda. In other words, individuals, driven by an inclination for conceptual coherence that relies on finding clusters of covarying attributes, may violate statistical principles to form a more coherent image of Linda based on the provided description.

A similar approach to the one we are proposing was developed by Siebel (2002), although he relies on Thagard's coherence model (see also, Trpin and Hartmann 2024). An

interesting aspect of Siebel's analysis is that it connects coherence (from the perspective of classical epistemology[6]) with inference and then with explanation. For Siebel, subjects are inclined to say that Linda is a feminist bank teller because this piece of information allows us to establish more 'inferential connections' with the previous information we have about her, something that positively impacts our ability to generate explanations of Linda's features. From our perspective, the link to explanation emerges naturally as inferential connections are reduced to (and explained in terms of) our knowledge of covariations embedded in the structure of the concepts we use.

Interestingly enough, there is experimental evidence on the cross-classification of people and induction that supports the previous ideas (Patalano et al. 2003, Patalano et al. 2006). We seem to have a clear preference for the use of high-coherence categories as a basis for the explanation of people's features, even if our reasoning apparently violates the laws of probability.

7. Conclusion

The article presents a novel analysis of the nature of conceptual coherence and the methods for measuring it. The central idea is that intuitions about coherence arise from identifying clusters of covarying properties within concepts (understood as sets of points in a conceptual space). These clusters can be synthesized into new summary dimensions that capture the intrinsic variability of those properties. Identifying these covariations can reduce the uncertainty associated with a concept's inherent variability. In other words, the perception of a concept as

[6] Siebel follows Bonjour in this regard: The coherence of a system of beliefs is diminished to the extent to which it is divided into subsystems of beliefs which are relatively unconnected to each other by inferential connections (Bonjour 1985, p. 98).

coherent or incoherent depends on how effectively our cognitive system compresses the information contained in it.

Our analysis also explores the relationship between coherence and induction, a connection consistently highlighted by experimental studies. This relationship is important because it underscores the fundamental role of coherence in shaping inductive reasoning, guiding predictions and generalizations based on category memberships across different concepts (natural, artifacts, and social).

Another connection worth exploring is that between coherence and the essences that people attribute to the categories that our concepts represent. In Gärdenfors and Osta Vélez (to appear), we analyze the notion of *essence* in terms of principal components. The basic idea is that the stronger the first principal component(s) are, the more willing subjects are to ascribe an essence to the concept (Gelman 2003). The essence of a concept is also seen as the cause of the cause of the inductive inferences drawn about it.

We demonstrate that the theory of conceptual spaces provides a robust framework for understanding conceptual coherence. While PCA highlights its potential as a measure of coherence, other methods, such as Mixed-Variable Factor Analysis and Multiple Correspondence Analysis, may be more suitable for concepts involving continuous and discrete variables. Future work could explore these methods to further refine our understanding of coherence across diverse conceptual structures, ensuring that the chosen approach fits the complexity and diversity of our conceptual thinking.

References

Ahn, W. K., Marsh, J. K., Luhmann, C. C., & Lee, K. (2002). Effect of theory-based feature correlations on typicality judgments. *Memory & Cognition*, 30(1), 107-118.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, 11(3), 211-227.

Billman, D., & Dávila, D. (2001). Consistent contrast aids concept learning. *Memory & Cognition*, 29(7), 1022-1035.

Billman, D., & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(2), 458-475.

BonJour, L. (1985). *The structure of empirical knowledge*. Harvard University Press.

Brandone, A. C. (2017). Changes in beliefs about category homogeneity and variability across childhood. *Child Development*, 88(3), 846-866.

Douven, I., & Gärdenfors, P. (2020). What are natural concepts? A design perspective. *Mind & Language*, 35(3), 313-334.

Douven, I., & Meijs, W. (2007). Measuring coherence. *Synthese*, 156(3), 405-425.

Douven, I., Verheyen, S., Elqayam, S., Gärdenfors, P., & Osta-Vélez, M. (2023). Similarity-based reasoning in conceptual spaces. *Frontiers in Psychology*, 14, Article 1130979.

Gärdenfors, P. (2000). *Conceptual spaces*. MIT Press.

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. MIT Press.

Gärdenfors, P., & Osta-Vélez, M. (2024). Generics as expectations: Typicality and diagnosticity. *Ratio*, 37(1), 5-23. https://doi.org/10.1111/rati.12424

Gärdenfors, P., & Osta-Vélez, M. (to appear). The essence of a concept is its principal components.

Gelman, R., & Williams, E. M. (1998). Enabling constraints for cognitive development and learning: Domain specificity and epigenesis. In W. Damon (Ed.), *Handbook of child psychology* (pp. 575-620). Wiley & Sons.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology*, 20(1), 65-95.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford University Press.

Giere, R. N. (2000). Theories. In W. H. Newton-Smith (Ed.), *A companion to the philosophy of science* (pp. 515-524). Wiley-Blackwell.

Goodman, N. (1983). *Fact, Fiction, and Forecast*. Harvard University Press.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. MIT Press.

Hampton, J. A. (2007). Typicality, graded membership, and vagueness. *Cognitive Science*, 31(3), 355-384.

Hayes, B. K., Taplin, J. E., & Munro, K. I. (1996). Prior knowledge and sensitivity to feature correlations in category acquisition. *Australian Journal of Psychology*, 48(1), 27-34.

Holland, J., Holyoak, K., Nisbett, R., & Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery*. MIT Press.

Johannesson, M. (2002). *Geometric models of similarity* [Doctoral dissertation]. Lund University.

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer.

Jones, E. L., & Ross, B. H. (2011). Classification versus inference learning contrasted with real-world categories. *Memory & Cognition*, 39(5), 764-777.

Kant, I. (1997). *Critique of pure reason* (P. Guyer & A. Woods, Trans.). Cambridge University Press. (Original work published 1781)

Kornblith, H. (1995). *Inductive inference and its natural ground*. MIT Press.

Lakoff, G. (2008). *Women, fire, and dangerous things*. University of Chicago Press.

Lassaline, M. E., & Murphy, G. L. (1996). Induction and category coherence. *Psychonomic Bulletin & Review*, 3(1), 95-99.

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

Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, 23(2), 250-269.

Medin, D. L., & Wattenmaker, W. D. (1987). Category cohesiveness, theories, and cognitive archeology. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 25-62). Cambridge University Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92(3), 289-316.

Nguyen, S. P., & Chevalier, T. (2015). Category coherence in children's inductive inferences with cross-classified entities. *Cognitive Development*, 35, 137-150.

Osta-Vélez, M., & Gärdenfors, P. (2022). Nonmonotonic reasoning, expectations orderings, and conceptual spaces. *Journal of Logic, Language and Information*, 31(3), 359-379.

Patalano, A. L., Chin-Parker, S., & Ross, B. H. (2003). The role of coherence in category-based explanation. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 910-915). Cognitive Science Society.

Patalano, A. L., Chin-Parker, S., & Ross, B. H. (2006). The importance of being coherent: Category coherence, cross-classification, and reasoning. *Journal of Memory and Language*, 54(3), 407-424.

Patalano, A. L., Wengrovitz, S. M., & Sharpes, K. M. (2009). The influence of category coherence on inference about cross-classified entities. *Memory & Cognition*, 37(1), 21-28.

Rehder, B., & Hastie, R. (2004). Category coherence and category-based property induction. *Cognition*, 91(2), 113-153.

Rehder, B., & Kim, S. (2006). How causal knowledge affects classification: A generative theory of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 659-683.

Rehder, B., & Kim, S. (2010). Causal status and coherence in causal-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1171-1206.

Rehder, B., & Ross, B. H. (2001). Abstract coherent categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(5), 1261-1275.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT Press.

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

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3), 382-439.

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.

Shipley, E. F. (1993). Categories, hierarchies, and induction. In D. L. Medin (Ed.), *Psychology of Learning and Motivation* (Vol. 30, pp. 265-301). Academic Press.

Siebel, M. (2002). There's something about Linda: Probability, coherence and rationality. In *First Salzburg Workshop on Paradigms of Cognition*. University of Salzburg.

Slaughter, V., & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. *Child Development*, 67(6), 2967-2988.

Sloman, S. A., & Lagnado, D. A. (2005). The problem of induction. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 95-116). Cambridge University Press.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, 7(6), 246-251.

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

Thagard, P. (2002). *Coherence in thought and action*. MIT Press.

Thagard, P., Eliasmith, C., Rusnock, P., & Shelley, C. P. (2002). Knowledge and coherence. In R. Elio (Ed.), *Common sense, reasoning, and rationality* (pp. 104-131). Oxford University Press.

Trpin, B., & Hartmann, S. (2024). Explaining the conjunction fallacy. In *Proceedings of the 46th Annual Meeting of the Cognitive Science Society*. Cognitive Science Society. https://escholarship.org/uc/item/4bv1k9nw

Turnbull, M. G. (2018). Underdetermination in science: What it is and why we should care. *Philosophy Compass*, 13(2), Article e12475.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327-352.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4), 293-315.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138(6), 1172-1217.

Yamauchi, T., Love, B. C., & Markman, A. B. (2002). Learning nonlinearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 585-593.

Younger, B. (1990). Infants' detection of correlations among feature categories. *Child Development*, 61(3), 614-620.