2-2021

Similarity Reimagined (with Implications for a Theory of Concepts)

Corinne L. Bloch-Mullins

Philosophy Faculty Research and Publications/College of Arts and Sciences

Theoria : A Swedish Journal of Philosophy, Vol. 87, No. 1 (February 2021): 31-68. DOI.

Similarity Reimagined (with Implications for a Theory of Concepts)

Corinne L. Bloch-Mullins
Marquette University

Abstract

Similarity-based theories of concepts have a broad intuitive appeal and have been successful in accounting for various phenomena related to the formation and application of concepts. Their adequacy as theories of concepts has been questioned, however, as similarity is often taken as too flexible, too unconstrained, to be explanatory of categorization. In this article, I propose an account of similarity that takes the "foil" against which the target items are measured as integral to the process of comparison, making the similarity relation a fundamentally triadic one. I argue that this account delivers more internal constraints on the process of comparison than the traditional, dyadic account of similarity does. Finally, I propose that this account is advantageous for a similarity-based theory of concepts, as it easily integrates with the notion of conceptual taxonomy and uses it to deliver additional constraints on categorization. On such a theory of concepts, the foil, whose variability is in large part responsible for the flexible nature of similarity judgements, is turned within the framework of conceptual taxonomy to the very element that constrains similarity judgements in categorization.

Therefore, this notion of similarity mitigates the worry that similarity is too flexible to be explanatory of categorization.

Keywords:

alignment; concepts; conceptual taxonomy; contrast classes; similarity

1. Introduction

Similarity-based theories of concepts maintain that we categorize items under the same concept in virtue of their similarity. The main similarity-based families of theories are prototype or exemplar theories. Versions of the prototype theory argue that a concept is a body of statistical knowledge about the features of category members (Posner and Keele, 1968; Rosch and Mervis, 1975). Versions of the exemplar theory maintain that it is a body of knowledge about the features of individual category members (Medin and Schaffer, 1978; Nosofsky, 1986, 1988).

While similarity-based theories have enjoyed great explanatory success with respect to some phenomena related to concept formation and application (for review, see Laurence and Margolis, 1999; Machery, 2009), critics have pointed out some explanatory shortcomings. A central difficulty stems from the fact that similarity is not absolute – it is always understood relative to a particular set of attributes (Quine, 1969; Goodman, 1972; Medin, 1989).[1] As Medin puts it, "a zebra and a barberpole would be more similar than a zebra and a horse if the feature 'striped' had sufficient weight" (Medin, 1989, p. 1473).[2] Consequently, it is often argued that similarity is too broad or unconstrained to be of explanatory value in a theory of concepts. Murphy and Medin, for example, point out that it is not enough to acknowledge that things in the same category are similar to each other. Without explicating the principles that determine the selection of the relevant dimensions used in the similarity judgements that give rise to categorization, the theory risks circularity: it may be that things appear similar simply because they belong in the same category (see Murphy and Medin, 1985, p. 291).

Various models of similarity acknowledge that similarity is always calculated with respect to a particular subset of attributes (see, for example, discussion in Decock and Douven, 2011), and that the selection of attributes that factor into similarity judgements changes with the context (Gärdenfors, 2000). The notion of context, however, is not well defined and may denote a broad range of factors such as goals, experience, environmental factors, etc. (see Machery, 2015). Thus, more work is needed for the similarity-based approach to account for the way in which different dimensions are selected, and the way in which they are weighed. Following Machery (2009), I refer to this as "the selection problem". [3]

One possible solution to the selection problem is to restrict similarity to perceptual features and suggest that similarity is therefore fixed and hard-wired (see discussion in Medin et al., 1993). This strategy has its appeals. As Goldstone (1994b, p. 136) points out, "it is difficult not to notice the similarity between a 400 Hz tone and a 402 Hz tone, or two shades of red". While some scholars seem to take similarity as calculated over purely perceptual features (e.g., Keil, 1986; Rips, 1989), advocates of the similarity-based approach tend to reject the idea that similarity judgements are so restricted. Eleanor Rosch and her colleagues, for example, take similarity to be calculated over both perceptual and functional attributes (Rosch et al., 1976a). Several psychologists and philosophers further argue that similarity can be calculated over abstract features or relations (Medin, 1989; Gentner and Rattermann, 1991; Medin et al., 1993; Goldstone, 1994b; Hahn and Chater, 1997; Hampton, 1998; Gärdenfors, 2000; Mazzone and Lalumera, 2010; Rheins, 2011; Bloch-Mullins, 2018). Moreover, Medin et al. (1993) argue that insisting that similarity is calculated over perceptual features would not necessarily provide firm grounding for similarity, as measures of perceptual similarity do not always converge on a single construct.[4] Finally, restricting similarity to purely perceptual aspects would drastically reduce the dominion of similarity and, therefore, its explanatory power for a theory of concepts (Medin et al., 1993, p. 256).[5]

It seems, then, that similarity-based theories of concepts are faced with a two-horned dilemma: if similarity is strictly perceptual, it cannot explain concepts that are grounded in non-perceptual features. If, on the other hand, similarity is not restricted to perceptual features, then the notion of similarity risks becoming circular. Moreover, even if the principles that govern the selection of relevant dimensions are fully elucidated, this would not necessarily vindicate the similarity-based approach to concepts. If factors such as goals or causal knowledge determine the relevant dimensions for similarity judgements, it may be argued that it is these factors that are doing the explanatory work, rather than the notion of similarity (see Medin, 1989, p. 1474). The burden is on proponents of the similarity-based approach, therefore, to show what is added to the explanation when we incorporate similarity into the account.

In light of criticism of the similarity-based approach, its popularity has reduced, and a number of different theories have been put forward. One prominent alternative is the theory-based approach, which maintains that a concept stores knowledge that is explanatory of the features of category members (Carey, 1985; Murphy and Medin, 1985; Gelman and Markman, 1986; Keil, 1989). For those who take similarity as computed over purely perceptual elements, the theory-based approach – which emphasizes the role of causal and abstract knowledge – is an appealing alternative. The theory-based approach has its own explanatory shortcomings, however, leading several authors to suggest, in recent years, that no "homogeneous" theory can account for all phenomena related to the formation and application of all concepts (Machery, 2005, 2009; Piccinini and Scott, 2006; Weiskopf, 2009). Alternative solutions included hybrid, pluralist and eliminativist theories. Hybrid theories of concepts maintain that a concept is comprised of several distinct parts, each part storing different types of information about the members of the category (for example, a theory part and an exemplar part) (Smith et al., 1974; Osherson and Smith, 1981; Anderson and Betz, 2001; Rice, 2016).[6] Pluralist theories propose that concepts do not belong to a single kind. Rather, there are different kinds of concepts (e.g., prototypes, exemplars, theories) (Machery, 2005, 2009, 2014; Piccinini and Scott, 2006; Weiskopf, 2009). Finally, eliminativist theories go even further and suggest that we should do away with the very concept of concept (Machery, 2009, 2014).[7]

I suggest that the flexible nature of similarity is not sufficient reason to disqualify the similarity-based approach to concepts, as similarity is not so flexible as to be nonexplanatory.[8] Specifically, I focus on a particular type of constraint, which originates from the process of comparison itself. Since this type of constraint is internal to similarity judgements, it suggests that the notion of similarity cannot be reduced to factors that are, in a sense, "external" to the comparison process, such as causal knowledge, goals, etc. Since the particular type of constraint I discuss does not depend on a grasp of purely perceptual similarities, my view advises optimism about the ability of the similarity-based approach to avoid the two-horned dilemma discussed above; that is, about the prospect that such a theory could deliver constraints on similarity – and, therefore, on categorization – without restricting its dominion to concepts that depend strictly on perceptual similarity.

In section 2, I discuss one specific way in which similarity is context-sensitive. My focus is on a particular element of context – the items in the background, against which the targets are compared. I refer to these items as the "foil".[9] I discuss different ways in which the foil affects similarity judgements, constraining them in a systematic manner. I then draw on works on alignment in comparison processes to reconstruct the notion of similarity, suggesting that the foil is best understood as a relatum in the similarity relationship.

Section 3 draws on work on categorization to discuss the way in which the notion of similarity I propose may be employed by a similarity-based theory of concepts. Specifically, on such an integrated view, a conceptual taxonomy, in which concepts are embedded within a nested hierarchy of categories, systematically fixes the elements that act as a foil in similarity judgements. Thus, the taxonomy itself constrains similarity judgements in categorization and facilitates their relative stability.

Section 4 proposes that the notion of similarity developed here helps mitigate the worry that similarity is so flexible and unconstrained that it cannot be explanatory in a theory of concepts. I argue that this account of similarity expands the internal constraints on similarity beyond those offered by traditional accounts. Moreover, expanding these constraints does not come at the cost of restricting similarity to the realm of the perceptual; similarity can be calculated over relational features at different levels of abstraction, including causal or abstract features. This suggests that a similarity-based theory of concepts can deliver constraints on categorization without restricting its scope to concepts close to the perceptual level.

2. Similarity as a Triadic Relation

2.1 Foil effects

In what follows, I assume a geometric framework for similarity, such as Gärdenfors's (2000) notion of conceptual spaces.[10] A conceptual space is a multi-dimensional space in which the dimensions represent various ways by which stimuli are judged as similar or different (e.g., 'weight', 'temperature').[11] A particular instance is represented by the coordinates of a point within the space (e.g., '5 kg', '100° C'). The degree of similarity of two objects is an exponentially decaying function of the distance between their representations, along the relevant dimensions (e.g., Hahn and Chater, 1997). The scales used for the determination of the distance along each dimension, however, are not static: the distance, in conceptual space, between the representations of any two particulars along the relevant dimensions may "stretch" or "shrink" (Hahn and Chater, 1997, p. 14; Gärdenfors, 2000, p. 20).[12] To allow for such stretching and shrinking, these accounts incorporate selective attention into the model. Each dimension (m) receives an "attention weight" (W_m), such that $0 \leq W_m$ and $\sum W_m = 1$. Increased attention weight for a given dimension leads to stretching of the scale along this dimension while shrinking it along others.

Allowing the scales to be dynamic is a way in which geometric models can accommodate the flexible nature of similarity. But, as discussed in section 1, it is not enough for a theory of similarity to acknowledge such flexibility; ideally, it should have something to say about the factors that affect it. In other words, a theory of similarity should specify how W_m values for the various dimensions are determined. Several studies show that factors such as goals, theories and background knowledge affect the centrality of dimensions in similarity judgements (e.g., Barsalou, 1982; Ahn et al., 2000). In what follows, I discuss an additional factor that influences the attention weight given to each dimension, namely, the foil, i.e., the other particulars taken into account.

I look at three systematic effects that the foil has on the centrality of dimensions: diagnosticity effects, range effects and density effects. An example of diagnosticity effects is illustrated in Figure 1.[13] Here, the similarity between the pair 'x' and 'y' is low when the set includes 'w' (Figure 1A) but increases when the same pair is compared within a set that includes 'z' (Figure 1B). In the first set, the most salient relevant dimension for the similarity judgement is 'brightness'; since all objects in the set share the same 'shape', the latter dimension carries relatively little weight. In the second set, however, the salience of the 'shape' dimension increases, leading to an increase of the similarity rating for 'x' and 'y'.

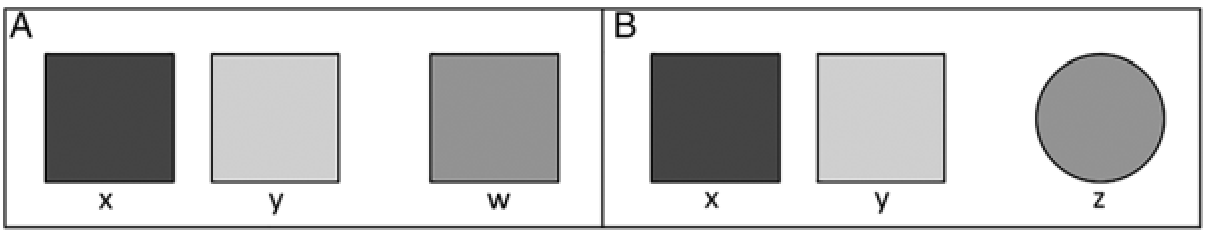

Fig. 1 Diagnosticity effects of the foil. Note: Variation in values along a dimension increases its centrality in similarity judgements.

Several studies show that dimensions weigh more heavily in similarity judgements when the particulars in the set exhibit some variation in values along these dimensions (Torgerson, 1965; Tversky, 1977; Tversky and Gati, 1978; Medin et al., 1993; Goldstone et al., 1997). For example, the average similarity rating for the pair of words "black" and "white" was 2.2 (on a scale of 1–9) when the two words were presented by themselves but increased to 4 when "red" was added to the set (Medin et al., 1993). The authors explain that:

When antonyms are compared by themselves, the single difference between them is highly salient. The properties that are shared by the antonyms fall into the background. The fact that white and black are at opposite ends of the gray scale is particularly salient. The fact that black and white are both monochrome colors on endpoints of the gray scale is not as salient. This commonality becomes more important when red is included in the context. (Medin et al., 1993, p. 268)

Importantly, these effects extend beyond perceptual features. For example, when subjects were presented with sets of countries, and asked to rate the similarity of pairs of countries within a set, the attributes 'American' and 'European' became more central in similarity ratings when they were not shared by all countries in the set (Tversky, 1977).

It should be noted, however, that the effects of the foil on the mapping of particulars onto conceptual space, that is, on the stretching and shrinking of the scale, are not restricted to selection among dimensions. This becomes apparent when we examine similarity judgements made with respect to a single dimension. The similarity between two points along a dimension depends on the overall range of stimuli. An example of range effects is provided in Figure 2: the similarity between the sizes of 'x' and 'y' seems low when the pair is contrasted with 'w' (Figure 2A) but increases when the same pair is contrasted with 'z' (Figure 2B). This intuitive judgment is supported by empirical evidence. For example, King and Atef-Vahid (1986) show that two small black disks, presented along with a large black disk, were judged as more similar when the size of the large disk increased. These results are in accordance with Krumhansl's (1978) distance–density model, which proposes that when the range of the stimuli is increased by adding more extreme stimuli along the relevant dimension, the similarity between two objects increases.

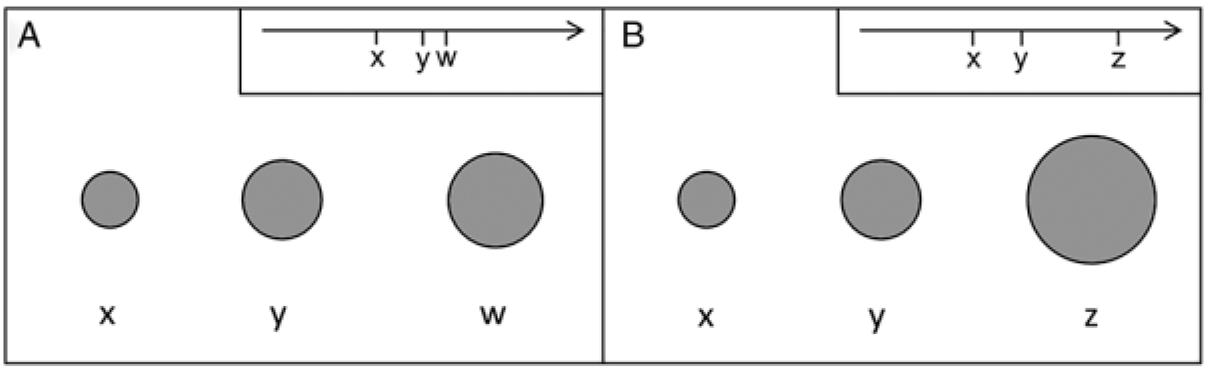

Fig. 2 Range effects of the foil. Notes: The foil affects similarity judgements via the alteration of the range of values of the objects in the set, along the relevant dimension ('diameter'). Insets: scales depict the relative diameters of objects x–z.

Additionally, according to the distance–density model, the similarity between two points along a dimension depends on the local density of stimuli along subregions within the set. The model proposes that the similarity between two objects decreases when they are located along a relatively dense subregion of the dimension in question and increases when they are located along a sparse subregion (Krumhansl, 1978; but see Corter, 1987). This density effect is illustrated in Figure 3, where the similarity between the sizes of 'x' and 'y' seems low when they are positioned in a dense subregion of the size scale (Figure 3A) but increases when the same objects are positioned within a sparse subregion of the size scale (Figure 3B).[14] If similarity judgements are not restricted to perceptual dimensions, we should expect to see these effects of the foil on similarity calculated over non-perceptual dimensions as well.

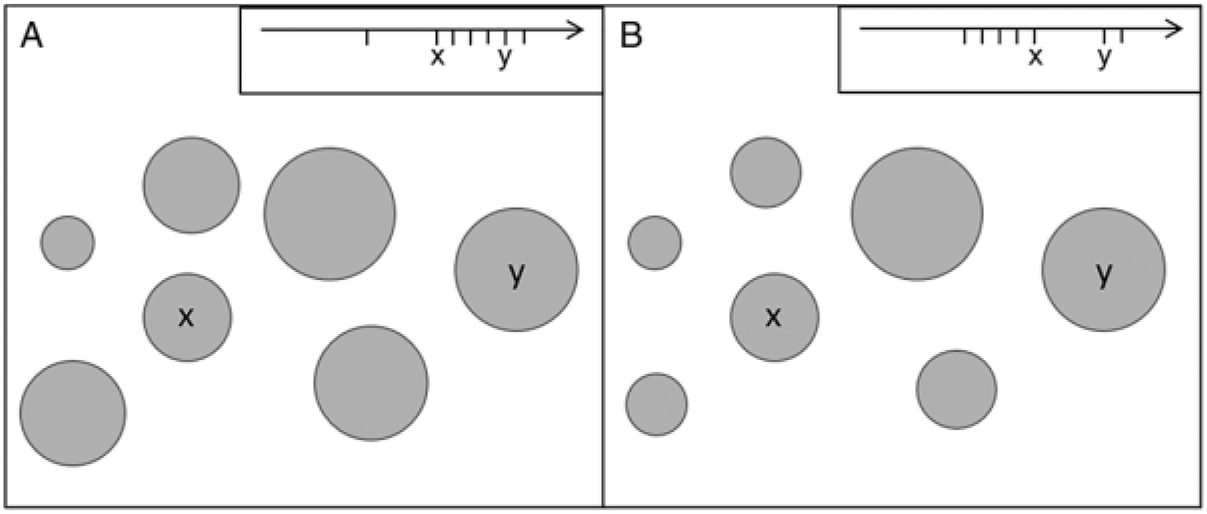

Fig. 3 Density effects of the foil. Notes: According to Krumhansl's (1978) model, the foil may affect similarity judgements via the alteration of local densities of the objects in the set, along the relevant dimension ('diameter'). Insets: scales depict the relative diameters of all objects in the set.

2.2 The ubiquity of the foil

As we have seen, the foil systematically affects similarity judgements by altering the relative centrality of different dimensions used in the comparison: all things being equal, dimensions gain weight when their values vary among members of the stimulus set. Moreover, the foil systematically affects the stretching and shrinking of the scale even when the comparison involves only a single dimension, via the determination of the overall range of values and, if Krumhansl's (1978) model is correct, the local densities of the stimuli in the set.

I would like to propose now that the foil is ubiquitously employed in all similarity comparisons. Consider the following example:

(i) The cost of living in Seattle is similar to the cost of living in New York City.

Although the relevant dimension – 'cost of living' – is specified,[15] it is impossible to evaluate statements such as (i) without using some standard for comparison, or points of reference. For example, the sentence may be judged as true when taking into consideration all cities worldwide, and as false when taking into consideration only major American cities. Similarity judgements require some points of reference, which are provided by the other particulars under consideration; that is – by the foil. I suggest that the foil is ubiquitously employed as a standard for comparison in such judgements.

My proposal may be illustrated further by drawing some parallels with discussions in linguistic semantics about the interpretation of relative gradable adjectives (e.g., "tall").[16] A common view is that the degree to which a particular possesses a relative gradable adjective is determined by the predicate and a standard of comparison that is context-dependent (see discussion in Kennedy, 2007). The applicability of the predicate "tall" to an individual is determined by having height that is at least as much as a standard of comparison of height. The value of this standard is not part of the lexical meaning of "tall" but is determined "on the fly" in accordance with the context. Whether or not the sentence "John is tall" is judged as true varies with the standard of comparison (Kennedy, 2007). This standard of comparison is computed relative to a comparison class (e.g., Klein, 1980).[17] The sentence may be true if the comparison class includes all people, but false if it only includes professional basketball players. Along the same lines, my suggestion is that the judgement whether the cost of living in Seattle is similar to the cost of living in New York City requires a comparison class, i.e., a foil. (Here, too, the foil is not part of the lexical meaning of "similar" but is determined on the fly as similarity judgements are executed.)

My position here, then, is not that the foil merely *modifies* similarity judgements, but that it is *required* for such judgements to take place. According to the former view, there is some sort of a "default similarity judgement", which is calculated over two items, and this judgement can then be modified when additional items (i.e., the items that serve as the foil) are presented. Goldstone (1994b, p. 139), for example, seems to endorse such an approach when he argues that while similarity is, to an extent, context-sensitive, there is a stable core that is relatively context-independent; a sort of automatic perception of "generic" similarity. In contrast, under the present proposal, similarity judgements must always incorporate a foil. We cannot judge the similarity between Seattle and New York City, with respect to cost of living, without drawing on some foil for a standard of comparison. I further propose that, given that the foil is ubiquitously employed in similarity judgements, that it can be teased out of the broad umbrella of "context", and that its effects can be systematized, it should be incorporated into a model of similarity.[18]

2.3 The foil as a relatum in similarity relations

I have argued that the foil is *ubiquitously employed* in similarity judgements, and that it behaves *systematically* and *consistently* in affecting both the centrality of dimensions in similarity judgements and the mapping of magnitudes onto scales in conceptual space. In what follows, I make an additional proposal, namely, that the foil is partially *constitutive* of the similarity relata. Accordingly, we should think of similarity as a triadic relation rather than a dyadic one.

Clarifying and defending this position requires that we shift our focus to the *process* of similarity judgements. Gentner and Medina distinguish between "similarity as a *process* of comparative reasoning and similarity as a *product* – e.g., a sense of closeness or representational unity" (Gentner and Medina, 1998, p. 266; original emphasis). While it is important to avoid ambiguity by distinguishing these two senses of "similarity", a naturalized approach to similarity must not disconnect the similarity relation, as a product, from an account of the process by which similarity is evaluated. Rather, it must take the process of comparison as a guide to understanding the sort of relation that similarity, as a product, evaluates. More specifically, it should view the objects being compared in the process that produces similarity judgements as comprising the relata of these judgements. To see how we can identify the objects that are compared during similarity judgements and distinguish them from the factors working in the background, more needs to be said about the process of comparison.

Medin et al. argue that the process of making similarity judgements is more like a dynamic search process than a computation over a fixed feature space. A crucial aspect of this process is alignment, that is, the procedure by which aspects of the entities being compared are brought into correspondence (Medin et al., 1993). We may understand alignment as the identification of corresponding attributes – or, using the terminology of the geometric approach, values that are located along the same dimensions in conceptual space – of the compared objects.[19]

The alignment process itself is not trivial, as there are different ways to carve out and align features. Medin et al. use a simple example to illustrate this complexity:

> Suppose that Person 1 has a striped shirt and that Person 2 has both a striped shirt and striped pants. If one allows *striped* to be a feature, does one count just one match or two? If one decides to count just one match, then would Person 3, with a plain shirt and striped pants, also have one match on the feature *striped* when compared with Person 1? If so, would the match count exactly the same as a striped shirt match? Now make the situation slightly more complicated. Person 1 has on a black and white striped shirt and red and green checkered pants, and Person 2 has on a black and white checkered shirt and red and green striped pants ... Is one allowed to count both the red and green matches and the striped and checkered matches, or does a commitment to one exclude the other? That is, if one aligns striped with striped, does one get a mismatch for colors? Implicit in this example is the idea that structure and global consistency, rather than simple local matches, may be important in the process of determining similarity. (Medin et al., 1993, p. 260, original emphasis)

The structure-alignment theory of similarity argues that people align features to produce a "structurally consistent" match by adhering to two specific principles: one-to-one correspondence and parallel connectivity (see Markman and Gentner, 1993b). One-to-one correspondence requires that individual elements in one representation are each placed in correspondence with at most one element in the other representation. Parallel connectivity requires that the arguments of corresponding predicates are themselves placed in correspondence. When more than one structurally consistent match is possible, people prefer the alignment that is most systematic, producing deeply connected relational matches rather than matching local elements and unconnected relations. The structure-alignment view, then, understands the process of similarity judgement as *a dynamic comparison process, in which the attributes and structures of the compared objects themselves affect the selection of the elements that get aligned along the same dimensions, in a way that maintains structural consistency and maximizes systematicity* (Markman and Gentner 1993b). This alignment then serves as the basis for the identification of commonalities and differences between the stimuli (Markman and Gentner, 1993a).

Several studies have shown that aligned features have a greater effect on similarity judgements than misaligned ones, indicating the importance of alignment in similarity judgements (e.g., Goldstone, 1994a; Markman and Gentner, 1996). As the above discussion makes clear, the determination of the elements that get aligned in the process is not fixed *a priori*; rather, it depends on the attributes of the objects being compared (Medin et al., 1993; Markman and Gentner, 1993a, 1993b, 1996; Gentner and Markman, 1994; Goldstone, 1994a). Here, I suggest, a triadic notion of similarity has an explanatory advantage over a dyadic one, as it enables us to understand the effect of the foil on the centrality of dimensions in similarity judgements from the perspective of the alignment-based approach: under this proposal, the objects whose features are aligned in the comparison process that generates similarity judgements are not only the target objects, but also those that comprise the foil.

Let us take Tversky's (1977) study as an example. He reported that similarity ratings for pairs of European countries (e.g., Italy–Switzerland) were higher when the stimuli set also included American countries than when the set included only European countries. If we understand the similarity comparison as a process in which the attributes and structures of the compared objects themselves affect the selection of the elements that get aligned to achieve structural consistency and maximize systematicity, then Tversky's results could be explained by the hypothesis that participants bring the attributes of *all* countries in the set into alignment, rather than only those of the pair that is the explicit target of the similarity judgement task. As part of the alignment process, the attributes 'American' and 'European' – which play the same role (or, put in terms of the geometric approach, can be represented as different values placed along the same dimension) – are brought into correspondence, adhering to the principles discussed above. The salience of the dimension 'belongs to continent x', as well as its weight in similarity judgments, increases due to the variation in values, among set members, along this dimension. This, in turn, increases the similarity between Italy and Switzerland, which share a value along this dimension. Importantly, it is the identification of 'American' and 'European' as different values on the same dimension (that is, as attributes that can be aligned) that increases the psychological availability of the relevant difference between Italy and Switzerland on the one hand, and Brazil and Uruguay on the other (I draw on Gentner and Markman, 1994; Gentner and Gunn, 2001). Taking the foil as integral to similarity judgements, then, allows the incorporation of the effect of the foil on the centrality of dimensions in similarity judgements into an alignment-based model of comparison.

In sections 2.1 and 2.2, I suggested that the foil is ubiquitously employed in similarity judgements, and that its effects are systematic. Here, I have taken another step, proposing that there is an explanatory advantage to viewing the foil as partially comprising the objects compared in the process of similarity judgements. This move enables us to understand foil effects not as a distinct phenomenon, but simply as arising from the same principles governing alignment during the comparison process: the attributes of the foil are aligned with the attributes of the targets; in this process, the attributes of *all* the compared objects influence (and constrain) alignment in an interactive, dynamic fashion aimed at achieving structural consistency and maximizing systematicity.

I further suggest that including the foil among the objects being compared has implications for the way we ought to understand the similarity relation itself. If we are committed to a naturalized approach to similarity in which the *process* of alignment in similarity judgements is intimately connected with the *outcome* of these judgements, we should view all the objects being compared in the process and whose attributes are aligned – and not merely a subset of these objects – as the substrates of similarity judgements. On this proposal, grasping 'x' and 'y' as similar amounts to grasping that 'x' and 'y' are closer to each other than they are to (an implicit or explicit) foil 'f', along a subset of dimensions.[20] This entails that not only the targets, but also the foil, comprise the relata of similarity judgements; the relation that is evaluated in similarity judgements is a triadic, rather than a dyadic, relation.[21]

My proposal that the similarity relation is best characterized as triadic is in line with Goldstone et al.'s (1997, p. 238) statement that "similarity is not just simply a relation between two objects; rather, it is a relation between two objects and a context" ("context" here is used narrowly, to refer to the set of particulars under consideration).[22] Along the same lines, Rheins (2011, p. 256) proposes that "Similarity consists in relatively little difference, where the difference between the compared units is small relative to their far greater differences from some other thing, the foil, in the same dimension(s) of variation".

Providing a complete argument for this position is outside the scope of the present article (in particular because it is partly dependent on open empirical questions).[23] I propose, however, that incorporating the foil into the similarity relation has explanatory, conceptual and methodological advantages. First, as discussed above, it enables us to use the resources of the structure-alignment view of similarity to account for foil effects, facilitating explanatory unification between studies examining the effects of the foil on the outcome of similarity judgements and models that focus on the process of comparison and the internal constraints that are placed on this process. Moreover, it formalizes a standard for similarity judgements while still allowing flexibility, given that the value of the standard can be determined on the fly during specific acts of judgement. Finally, taking similarity as fundamentally triadic has implications for further empirical research, as it prompts the exploration of alignment of foil and target items during the comparison process. Rather than taking the foil as noise that needs to be controlled, my approach encourages psychologists to bring the foil to the forefront, as a variable that can provide important insight about the process of comparison and the judgements that it produces.[24]

2.4 Objections

In what follows, I examine two possible objections to my proposal. Both maintain that the foil is not strictly required, but on different grounds.

Objection 1: It might be pointed out that it is not meaningless to ask how similar two items are, without any explicit reference to a foil. Moreover, people are clearly successful in performing similarity judgements without such reference. For example, one can make judgements such as:

(ii) John and Jack have similar heights.[25]

without any explicit reference to a standard of comparison. I do not deny here that one can – and often does – perform such judgements. I suggest, however, that in such cases a foil is implicitly assumed. As Medin et al. (1995) point out, similarity comparisons are made in relation to an implied or constructed contrast set, or to some ideal drawing on prior examples (see also Rheins, 2011, p. 282, n. 5).

Here, again, work on relative gradable adjectives provides a useful illustration: the comparison class that determines the standard for the application of such an adjective may be unspecified; yet it is assumed implicitly. For example, the sentence "Fido is smart", when 'Fido' designates a dog, is judged against a standard that is derived from the general population of dogs unless otherwise stated (e.g., Klein, 1980).

To return to the notion of similarity, the present suggestion is that when one is tasked with evaluating statements such as (i) and (ii), but is not provided with a foil, one would implicitly draw on *some* foil as basis for comparison.[26] Let us suppose, for example, that John is 6’3” tall and Jack is 6’5” tall. If all we know is that both John and Jack are adult American men, we would probably evaluate the similarity between their heights in relation to what we know about the general distribution and range of heights among adult American men. In this case, we are likely to make the judgement that their heights are fairly similar. If, on the other hand, we know that both John and Jack are basketball players, we may implicitly evaluate the similarity between their heights in relation to what we know about the general distribution and range of heights among basketball players. In this case, we are more likely to make the judgement that their heights are quite dissimilar.

Thus, I suggest that the foil is not something that merely modifies, or acts upon, similarity judgements. Rather, it both gives rise to – and constrains – these judgements. We can only grasp Seattle and New York City as similar with respect to 'cost of living' when we compare them to other locations (or some subset of locations). We can only grasp John and Jack as similar with respect to 'height' when we compare them to other individuals (or some subset of individuals). The present suggestion, then, is that *there can be no similarity judgements without a foil*, whether implicit or explicit.

Objection 2: One could object that the role of the foil is indirect and as such it is not strictly required. Accordingly, one might concede that the studies discussed above show that the foil is used for the determination of the relevant dimensions for similarity judgements, as well as an average value, or a range of values, along the dimension(s) deemed relevant. If, however, this information is provided via other means, then one could perform similarity judgements without using an explicit or implicit foil. Therefore, the objection goes, the foil should not be taken as a relatum in the similarity relation.

As a first pass, I note that information about the relevant dimensions for similarity judgements, as well as the average values and/or range of values on the scale of the relevant dimensions, might not be sufficient to uniquely determine similarity. This is illustrated in Figure 4, which represents values along a specific dimension – the 'cost of living' index – for two sets of cities. The two sets contain the same number of cities; they share the same range of values for 'cost of living' as well as the same average value. Note also that the 'cost of living' index values for Seattle and New York City are identical in both sets. However, when asked to judge the similarity between Seattle and New York City, with respect to 'cost of living', it is not obvious that we would give the same rating in both sets. Recall, again, Krumhansl's (1978) proposal that similarity judgements are affected by local density of the stimulus set along the range of the relevant dimension. If this is correct, then we would expect that the similarity between Seattle and New York City, with respect to 'cost of living', would vary among the two sets even though the sets exhibit the same range of values and the same average value.[27]

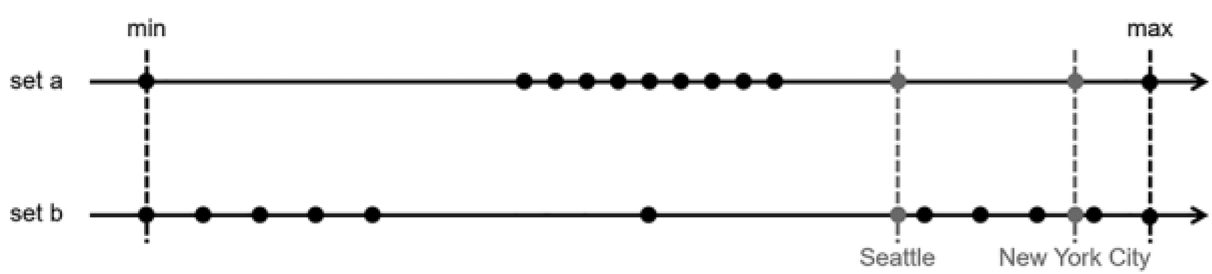

Fig. 4 Cost of living comparison for two sets of cities. Notes: For each set, points represent values for individual cities along a scale of 'cost of living' index. Note that the scale is presented on a theoretical, rather than a phenomenal, dimension (see n. 12). The figure therefore does not convey the stretching or shrinking of the scale that might take place in conceptual space, affecting one's similarity judgements with respect to the pair Seattle–New York City within each of the sets.

This offers some preliminary grounds, I think, to suspect that the foil provides rich information that is used in the process of similarity judgements; information that cannot be reduced to a specification of the relevant dimensions and the range and average of values along these dimensions. Again, none of this is to deny that it is perfectly coherent to make similarity judgements about the cost of living in Seattle and the cost of living in New York City without concrete knowledge about the local distribution of particulars in the relevant comparison class. The present suggestion is just that, when such information is not provided, it is implicitly assumed.

As I argued in the previous subsection, however, there is a more fundamental reason to hold that the foil is integral to similarity judgements. If we understand the process of comparison as one that begins with the alignment of the objects being compared, and if we acknowledge that the items that serve as the foil are aligned – along with the targets – as part of that process, then a complete account of the similarity relation should incorporate the foil objects as partially constituting its relata. An account that excludes the foil by beginning after the relevant dimensions and the granularity of the scale are determined gives us only a part of the story about the process of producing similarity judgements. Moreover, for an account of similarity to properly reflect the flexibility inherent in similarity, it must take some of its context-dependency as intrinsic to similarity calculations (Sloman and Rips, 1998). An account that takes the determination of the relevant dimensions and scales out of the realm of similarity computations risks losing some of the fundamental features of similarity, by treating it as a rigid construct that is calculated over a static set of features by determining distances on a fixed scale. Thus, I suggest that an account that begins with the process of comparison by which attributes are aligned and relevant dimensions for similarity judgements are selected better fits a view of similarity judgements as flexible and dynamic search processes.

Much of the above discussion highlighted the flexible, context-dependent, dynamic nature of similarity. It may seem as if this plays into the worry that motivated this discussion in the first place: the worry that similarity is too flexible, and thus cannot place constraints on concept formation and application. As Medin et al. (1993, p. 272) point out, however, the dynamic and context-dependent character of similarity is not a difficulty for a theory of concepts, if similarity is *systematically fixed* in context. In this section, I proposed that an understanding of similarity as a triadic relation that takes the foil as a relatum enables us to understand the systematic effects of the foil on similarity judgements, from the perspective of theories of alignment, as *internal constraints on similarity.* With this notion of similarity in hand, I proceed, in the next section, to suggest a way in which the foil – and, therefore, similarity judgements – is systematically constrained within the context of a hierarchical conceptual structure.

3. Contrast Classes as Constraining Similarity Judgements in Categorization

3.1 Conceptual taxonomy

To see how the notion of similarity, as constructed above, can be incorporated into a similarity-based theory of concepts, and help alleviate worries about lack of constraints on categorization, we first have to acknowledge that concepts are not formed in a vacuum; they are formed as part of a conceptual taxonomy. Rosch et al. write:

A taxonomy is a system by which categories are related to another by means of class inclusion. The greater the inclusiveness of a category within a taxonomy, the higher the level of abstraction. Each category within a taxonomy is entirely included within one other category (unless it is the highest level category) but is not exhaustive of that more inclusive category (see Kay, 1971). Thus, the term level of abstraction within a taxonomy refers to a particular level of inclusiveness. A familiar taxonomy is the Linnean system for the classification of animals. (Rosch et al., 1976a, p. 383)

Within a given inclusive category (e.g., rodent), there are several categories at the same level of abstraction (e.g., chipmunk, squirrel, groundhog).[28] These categories form a "contrast set" (Kay, 1971). Along the same lines, I use the term "contrast classes" with respect to a given category to refer to the other categories *within the same contrast set.* For example, the contrast classes for chipmunk are squirrel, groundhog, etc.

3.2 Contrast classes and the constraints on similarity

In section 2, I argued that the foil comprises one of the relata in the similarity relation. In this section, I suggest that a category's contrast classes serve as a relatively stable foil for similarity judgements in categorization. The classification of Alvin as a chipmunk, for example, takes place when the judgement about Alvin's similarity to other members of chipmunk is performed against the background of members of contrast classes such as squirrel, groundhog, etc.[29]

It is helpful to introduce my view by contrasting it with the classical works by Rosch and her colleagues, who also emphasize that categorization depends on both similarity to category members and dissimilarity to contrast classes (e.g., Rosch, 1978, pp. 30–31). Rosch and Mervis (1975), for example, show that attributes with high "cue validity" – a measure that is defined for a given attribute in terms of its frequency among members of a category relative to its frequency among members of contrasting categories – are more central in categorization decisions than attributes with low cue validity. Accordingly, targets are classified faster as members of a category and ranked as more typical category members when they possess attributes that are not also possessed by members of contrast classes.

Rosch and her colleagues, however, do not take members of contrast classes to contribute to the similarity between category members; similarity is understood as a relation that holds only between the targets. Similarity between members of natural semantic categories, for example, is defined in terms of the overlapping of attributes (Rosch and Mervis, 1975, p. 575); similarity between strings of symbols is quantified through the number of shared letters (Rosch et al., 1976b); similarity between two shapes is measured via the overlap of the two juxtaposed (Rosch et al., 1976a). Rosch and her colleagues thus take the differences between categories within a contrast set as feeding, *alongside* similarity judgements, into typicality judgements and consequently into categorization decisions. This process results in cuts between categories that maximize both intra-categories resemblance and inter-category difference (Rosch and Mervis, 1975; Rosch, 1978).

The present view, on the other hand, takes contrast classes – in their role as foils – as partially constituting the similarity relata that give rise to categorization. Since all the objects whose attributes are aligned during a similarity comparison affect the alignment process and the resulting similarity judgements, the identification of commonalities between category members is not independent of the identification of differences between members of contrasting classes; rather, the identification of these respective commonalities and differences results from the *same alignment process, along the same dimension(s).*[30]

Several studies have shown that alignment of members of contrast sets is important for assessment of both between-group differences and within-group commonalities. Markman and Wisniewski (1997) examined whether dissimilarities between members of contrast classes were alignable (related to the commonalities) or non-alignable (unrelated to the commonalities). They asked participants to list commonalities and differences for pairs of categories (e.g., watch–necklace). The listed differences counted as alignable if participants made explicit or implicit mention of a different value along the same dimension for both objects. For example, a statement such as "A watch is worn on the wrist, and a necklace is worn on the neck" counted as an explicit mention of an alignable difference, and a statement such as "A watch and a necklace are worn on different parts of the body" counted as an implicit mention of an alignable difference. All other differences, such as "A watch has a face, but a necklace does not" counted as non-alignable differences (Markman and Wisniewski, 1997, p. 60). Unsurprisingly, pairs of basic level categories[31] that belonged in the same superordinate category (e.g., bed–couch from the superordinate category furniture; horse–cow from the superordinate category animals, etc.) received significantly more commonalities than pairs that belonged to different superordinate categories (e.g., fork–newspaper from the categories kitchenutensils and readingmaterial, respectively). Interestingly, pairs of basic level categories that belonged in the same superordinate category also had a significantly higher ratio of alignable-to-non-alignable differences, compared to pairs that belonged in different superordinate categories. Markman and Wisniewski's data highlight the importance of alignable differences for distinguishing between basic level categories within the same contrast set. Given that alignment increases the psychological availability of differences (e.g., Gentner and Gunn, 2001), these results suggest that conceptual taxonomy is structured such that members of contrast sets are aligned along the same dimensions, and that such alignment facilitates the identification of differences between categories within the set.

Additional insight comes from studies in which subjects learned artificial categories. Lassaline and Murphy (1998) compared the effects of alignable and non-alignable commonalities within and between artificially constructed categories on the ease of category learning.[32] Commonalities were taken as aligned if they occurred on the same dimension (two birds that display the pattern 'spotted' on the same body part exhibit an aligned commonality) and as non-aligned if they occurred on different dimensions (two birds that display the pattern 'spotted', each on a different body part, exhibit a non-aligned commonality).[33] Lassaline and Murphy showed that within a category, commonalities that are aligned facilitate learning more than non-aligned commonalities do. Across category boundaries, aligned commonalities hurt learning more than non-aligned commonalities. The authors point out that if we assume that categorization follows similarity judgements, then these results are in line with prior work (Goldstone, 1994a; Markman and Gentner, 1996) showing that aligned commonalities and differences have a greater effect on similarity judgements than misaligned commonalities and differences do. Lassaline and Murphy's results suggest that, during category learning, subjects align not only objects that belong to the same category, but also objects that belong to contrasting categories, and that these alignment processes facilitate the identification of within-category similarities and between-category differences.

My suggestion, above, was that within-category commonalities are not independent of between-category differences, as members of a category and members of contrasting classes are aligned for comparison along the same dimensions. We have seen that members of a category are aligned during within-category similarity judgements (Lassaline and Murphy, 1998). Moreover, the identification of differences between contrast classes depends on the commonalities among them; that is, members of contrasting classes are aligned along the same dimensions, facilitating the identification of alignable differences between these classes (Markman and Wisniewski, 1997; Lassaline and Murphy, 1998; Gentner and Gunn, 2001). These results are in line with my suggestion. More needs to be said, however, to show how aligned commonalities within categories are related to aligned differences across categories. In other words, I still need to establish that the identification of these respective commonalities and differences results from the same alignment process, along the same dimension(s).

Few studies have directly examined the relation, in categorization, between aligned within-category commonalities and aligned between-category differences. Billman and Dávila (2001) examined the ease of category learning under "consistent contrast" and "inconsistent contrast" conditions. Consistent contrast is manifested when within-category commonalities are aligned with between-category differences. For example, if members of all categories in the set are objects that make sounds and move around, but members of each category are distinguished from members of other categories by the specific type of sound they make and the specific type of movement they display, then the contrast is consistent. Inconsistent contrast is manifested when each category is distinguished by a different set of dimensions. For example, if members of category X are distinguished from members of other categories by displaying certain sounds and movements, while members of category Y are distinguished by displaying certain colours and mating patterns, then the contrast is inconsistent. Billman and Dávila showed that it is easier to learn categories under consistent contrast than under inconsistent contrast. In other words, learning was more efficient when the commonalities within each of the categories and the differences between categories were aligned along the same dimensions. Participants in the consistent contrast condition also performed better in generalization and induction tasks and were more likely to learn correlations between values of different dimensions, compared to participants in the inconsistent contrast condition. The authors propose that our learning procedures are biased to develop conceptual systems that exhibit consistent contrast, as categories embedded in such systems are more useful.

Taken together, the results discussed in this subsection suggest that people align members of contrasting categories and use these alignments to identify within-category commonalities and between-category differences. These findings are just what we would expect if contrast classes are brought into alignment during the comparison process and serve as the foil for similarity judgements.

We are now in a position to see how dynamic, context-sensitive similarity judgements may give rise to relatively stable concepts. If members of contrast classes are brought into alignment and serve as the foil for similarity judgements in categorization, then *our conceptual taxonomy fixes the foils* for similarity judgements in categorization, and thus stabilizes these judgements, producing the systematic effects discussed in section 2. This accords with data on the centrality of diagnostic features – that is, features that usefully distinguish members of the target category from members of its contrast class(es) – in categorization.[34] I have already mentioned Rosch and Mervis (1975), who found that subjects judge items with high total cue validity as good examples of a category, and that attributes with high cue validity are more central in categorization decisions than attributes with low cue validity. Additional data indicate that the effects of diagnosticity on the centrality of dimensions in categorization are mediated via similarity judgements. For example, Hsu et al. (2014) manipulated the diagnosticity of colour as subjects learned various categories. They show that subjects who learn categories for which colour is a diagnostic dimension produce higher similarity ratings for same-coloured object pairs than do subjects who learn categories for which colour is a non-diagnostic dimension.

Finally, several works show that diagnostic features hold a privileged status in a concept's representation (e.g., Cree et al., 2006; Hsu et al., 2014), thus lending further support to the suggestion that contrast classes' effects are fixed by conceptual taxonomy. Hsu and her colleagues (2014) report data indicating that the privileged status of a concept's diagnostic features is underwritten by automatic recruitment of these features when the concept is evoked, even during tasks for which this information is irrelevant. If we take the comparison process as involving contrast classes, then Hsu et al.'s proposal that contrasting categories – via modification of the diagnosticity of specific dimensions – affect a category's representation fits nicely with works on alignment that show that the comparison process itself leads to a change in a category's representation (see discussions in Lassaline and Murphy, 1998; Medin et al., 1993). Taken together, these works provide further support for the proposal that the comparison process that generates similarity judgements about category members (i.e., the targets) involves the alignment not only of these members, but also of the members of contrasting classes, and that these contrasting classes provide stability in future similarity judgements.

To summarize, the alignment of objects both within and across categories along the same dimensions – considered in light of the effects of the foil on similarity judgements discussed in section 2 – suggests that within-category similarity is not evaluated independently of between-category differences; rather, the location of the foil along the aligned dimensions is incorporated into judgements about similarity among category members. The task-independent privileged status of diagnostic features in a concept's representation further suggests a path through which a relatively stable set of contrast classes can place constraints on similarity judgements in categorization. Understood in this way, we can see that the foil, whose variability is in large part responsible for the flexible nature of similarity judgements, is turned within the framework of conceptual taxonomy to the very element that fixes – and therefore constrains – similarity judgements in categorization.

3.3 Worries about circularity

I have argued that when the present notion of similarity is incorporated into a similarity-based theory of concepts we may take contrast classes – in their role as foils – as incorporated into the relata that give rise to similarity judgements about category members. Accordingly, the classification of Alvin as a chipmunk takes place when the judgement about Alvin's similarity to other members of chipmunk is performed against the background of members of contrast classes such as squirrel, groundhog, etc. One might worry, however, that incorporating this view of similarity into a similarity-based theory of concepts leads to circularity: maintaining that contrast classes serve as a foil in similarity judgements seems to entail that similarity depends on extant categories. At the same time, similarity-based theories of concepts hold that the taxonomy itself is constituted by similarity relations. To avoid circularity, I need to show that my account allows for the initial formation of chipmunk in a way that does not depend on the *prior* conceptualization of chipmunk, squirrel and groundhog.

I should first clarify that it is not my claim here that one must explicitly establish contrast classes before making similarity judgements.[35] Rather, I have suggested only that a foil is brought into the comparison when such judgements are executed. If the relevant contrast set has already been established, as in the example of the classification of Alvin under chipmunk, contrast classes serve as the foil for similarity judgements with respect to the category in question, as discussed above, bringing to the forefront the dimensions that vary across the contrast set. What if a contrast set has not yet been established? In what follows, I discuss the process by which similarity judgements, as characterized above, can get off the ground and lead to the formation of categories without prior partitioning of the particulars into contrast classes.

As a first step, it may be helpful to consider data from free-sorting tasks, in which participants are presented with a collection of items and are asked to group them into categories as they see fit. Such studies can tell us whether alignment is involved in the initial formation of taxonomies. Work reporting data from free-sorting tasks using relatively simple stimuli structure often show that people choose to sort stimuli by using a single dimension, largely ignoring overall correlations (e.g., Medin et al., 1987). However, studies employing richer stimuli – which are presumably closer, in this sense, to the objects we classify in real life – provide a more complex image of free-sorting strategies. Billman and Davies (2005), for example, show that participants sorting complex objects prefer constructing taxonomic systems in which within-category commonalities and between-category differences were aligned. Clapper (2017) also shows that alignability is more important in free-sort categorization than commonalities of local elements. These results suggest that when subjects sort complex stimuli, the resulting taxonomy mirrors the similarity relationship, discussed above, among contrast classes in conceptual hierarchy.

How would such classifications get off the ground? In accordance with the above discussion, the present suggestion is that the process by which all objects are aligned constrains the similarity judgements that can give rise to the formation of a taxonomy. First, the alignment process is constrained by the requirement of structural consistency and maximization of systematicity (Markman and Gentner, 1993b) among the attributes of *all* objects in the set.[36] Given that alignment increases the psychological availability of differences (Gentner and Gunn, 2001), alignment of attributes of all the items in the set facilitates the grasp of aligned differences among the objects, increasing the weight of these dimensions in similarity judgements while reducing the weight of dimensions that receive close values across the stimuli set (see Torgerson, 1965; Tversky, 1977; Tversky and Gati, 1978; Medin et al., 1993; Goldstone et al., 1997).[37] Once the relevant dimensions are determined, the values exhibited by particulars, along these dimensions, are compared, and similarity judgements are executed. There is no requirement that a foil be designated *prior* to these judgements; rather, it emerges during the similarity comparison process, constrained by the process of alignment that determines the relevant dimensions. To put the same point in terms of what a judgement of similarity entails, the central claim of the previous section was that a judgement of similarity amounts to grasping two (or more) particulars as closer to each other, along the relevant dimensions, than to a third. Such a grasp does not require that "the third" be assigned in advance. A simple example is shown in Figure 5A. Here, it is immediately clear that the relevant dimension for comparison is the fill colour. Without prior designation of a foil, we grasp the similarity between objects 'x' and 'z'. This amounts to a grasp of a triadic relation: on the dimension of fill colour, 'x' and 'z' are closer to each other than they are to 'y'.

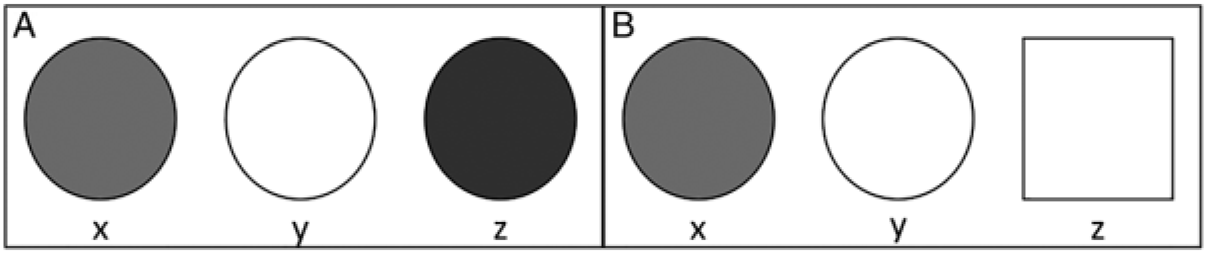

Fig. 5 Sets give rise to similarity relations without prior designation of foil items

Such simple similarity relations would not be sufficient for concept-formation (otherwise, *every* similarity judgement would entail a new concept). Rosch's (1978, p. 31) working assumption, for example, is that "(1) in the perceived world, information-rich bundles of perceptual and functional attributes occur that form natural discontinuities, and that (2) basic cuts in categorization are made at these discontinuities".[38] While the example in Figure 5A is rather impoverished, it illustrates, in a highly simplified form, how similarity judgements that could potentially lead to concept-formation can be constrained by the alignment process and get off the ground without any prior partitioning that designates foil items.

Of course, alignment may not be enough to constrain the way a given set may be partitioned. We can easily imagine a set in which objects vary along several dimensions, each diagnostic dimension giving rise to a different partitioning of the set. A simple example is depicted in Figure 5B. Here, aligned differences and commonalities still constrain the dimensions we take into account in similarity judgements: we do not take note of dimensions such as the colour or thickness of the objects' outline or the size of the objects; these dimensions do not vary among objects in the set, and therefore receive low weight in similarity judgements. The relevant dimensions here are 'shape' and 'colour'. However, these dimensions leave two alternative ways to partition the set.

Situations like that described in Figure 5B may be less likely to occur once we move from artificial stimuli to the rich stimuli we encounter outside the lab and that are actually used in the formation of real concepts – the sort of stimuli that Rosch has in mind. It should be pointed out, however, that even in the case of very rich objects, alignment and contrast effects may still not be enough to fully constrain similarity judgements that give rise to categorization. For example, attempts by early numerical taxonomists to classify organisms based on theory-free similarity did not result in the formation of stable categories (see discussion in Hull, 1988; Griffiths, 1997, p. 178). Several studies show that factors such as causal theories affect not only the attributes we use in categorization (Ahn et al., 2000), but also the correlations we take note of (see, e.g., Murphy and Medin, 1985; Ahn et al., 2002). My point in this subsection was not that alignment and contrast *fully* account for the formation of taxonomies. (Nor do I hold that a similarity-based account of concepts should be committed to some notion of theory-free similarity that would give rise to taxonomy in a fully determinate manner, and independently of other factors.) Rather, I defend the more modest claim that the foil constrains the process in the ways discussed above, and that these constraints – namely, the effect of contrast on alignment and on the centrality of dimensions in similarity judgements – need not depend on prior partitioning of the set, such that a foil (or a contrast class) is designated *before* similarity judgements can be executed.

My account thus avoids circularity as follows: when no prior partitioning of the objects is in place, the requirement to align all the objects in a way that satisfies structural consistency and maximizes systematicity places some constraints on the items that emerge as foil in similarity judgements. The process results in contrast sets in which within-category commonalities are aligned with between-category differences. Once categories are formed, the contrast classes embedded in a taxonomy help preserve stability in future similarity judgements – and hence the stability of categories – by regimenting foils that remain fairly constant across usages, as discussed in section 3.2.[39]

3.4 Context and concepts

Before concluding this section, a brief note is in order about the way in which the present position relates to the debate over the context-dependency of concepts ("context" here is construed broadly, to include a variety of factors). Invariantists take concepts to be context-independent representations (e.g., Keil, 1994; Mazzone and Lalumera, 2010; Machery, 2015). Contextualists, in contrast, take concepts to be context-dependent representations, constructed on the fly (e.g., Barsalou, 1993; Casasanto and Lupyan, 2015; Lebois et al., 2015; Yee and Thompson-Schill, 2016). It should be noted, however, that both invariantists and contextualists generally agree that (i) concepts change with time, experience, expertise, etc., (ii) context influences conceptual processing, (iii) there is some overlap between the representations retrieved when a given term is used across contexts, and (iv) the use of concepts is associated with some stable behaviours in the same individual and across individuals (see discussion in Löhr, 2017). While invariantists take the overlap in representations and behavioural stability to arise from the context-independent elements that constitute concepts (Machery, 2015), contextualists appeal to the elements of context that overlap among instances of conceptualization, such as language, culture, bodily elements, etc., thus giving rise to (seemingly) similar ad hoc constructs (Casasanto and Lupyan, 2015; Yee and Thompson-Schill, 2016). Consequently, the same empirical phenomenon may be accommodated by both views (see discussion in Bloch-Mullins, 2015; Löhr, 2017).

While each of these camps makes different commitments with respect to the various steps involved in conceptual processing, some of the debate between these camps can be understood as pertaining to the way in which concepts should be individuated and the extent to which they can be distinguished from context. Given the view of similarity I advance here and the role of the foil as a relatum in similarity judgements on the one hand, and as an element that is part of the broader context on the other hand (see n. 21), the present article is sympathetic to Yee and Thompson-Schill's (2016, p. 1016) statement that concepts are "inextricably linked to the contexts in which they appear, so much so that the dividing line between a concept and a context may be impossible to clearly make out". If one takes concepts as grounded in similarity relations and adopts the view of similarity proposed here, then one must acknowledge that similarity between category members is not context-free; rather, it depends (among other things) on the objects that serve as the foil. As Yee and Thompson-Schill (2016, p. 1019) point out, there is never an activation of a concept with no context. To the extent that the foil is entrenched within the conceptual taxonomy, however, part of the context is fixed across tasks and uses, giving rise to overlap in representations and behavioural stability among instances of conceptualization. So, it seems, we are back where we started: whether or not we take these stable elements as constitutive of concepts or as distinct from them depends on how we individuate concepts. This is a topic for another article.

4. Similarity as an Explanatory Construct in a Theory of Concepts

4.1 Constraints on similarity

The article proposes a broad notion of similarity that retains the virtue of flexibility by incorporating context-dependency but, at the same time, is less vulnerable to the worry that similarity is too unconstrained to be explanatory of concept formation and application. Such a project does not require that we establish that *all* constraints on categorization are internal to similarity judgements. For the notion of similarity to be taken as explanatorily valuable for a similarity-based theory of concepts, it must be shown there are *some internal constraints* on the similarity judgements that are intrinsically involved in categorization; that is, that there are *some constraints that arise from the comparison process itself.*

The studies discussed above, by Gentner, Markman, Medin, and their colleagues, have shown that the process of comparison by which similarity judgements are generated places internal constraints on the resultant judgement: the attributes of the aligned objects affect the global alignment that is then used as a basis for similarity judgements (see section 2.3). The structure-alignment theory of similarity argues that people align stimuli to produce a structurally consistent match by satisfying two principles: one-to-one correspondence and parallel connectivity. Moreover, mappings that involve deeply connected relational matches are preferred to correspondences involving only isolated, scattered matches. The resulting alignment serves as the basis for the identification of commonalities and differences between the stimuli (Markman and Gentner, 1993a, 1993b, 1996). Goldstone (1994a) extends these studies by providing a quantitative model of similarity judgments. In his model, correspondences and similarity are determined through a process of interactive activation in which feature, object and role correspondences mutually and simultaneously influence each other. That is, alignments of individual elements are not computed independently; rather, they are influenced by the general pattern of other alignments. We can see, then, how the principles governing the non-trivial process of alignment place internal constraints on the resulting similarity judgements. These constraints make similarity irreducible to other factors. Goldstone writes:

> There appear to be regularities concerning the salience of feature matches that cannot be attributable to influences of knowledge or individual featural saliences. Instead, to know whether a feature match between two scenes will count as a feature match (and how much it will count) for increasing similarity, it is necessary to know whether the feature match belongs to corresponding parts. (Goldstone, 1994a, p. 27)

The notion of similarity proposed here does not offer a *new type* of constraints on the process of comparison, beyond those offered by current alignment-based models. Rather, my suggestion in section 2.3 that we may understand foil effects (such as those reported in Tversky, 1977) from the perspective of the alignment-based approach amounts to a proposal to expand the constraints that theory had already put in place: if attributes of foil objects are aligned alongside attributes of target objects, then the relational and local features of the foil objects influence the alignment in the same way that the relational and local features of the target objects do. This alignment leads to identification of commonalities and differences among the items in the set, and diagnostic dimensions become more central in similarity judgements.

Similarity is a flexible, highly context-dependent construct. However, the dynamic, interactive process of comparison in which the targets and the foil are aligned places some internal constraints on the process. Thus, the context-dependency of similarity, at least to the extent that it arises from foil effects, is not an obstacle for a similarity-based theory of concepts, because it does not make similarity unruly and unconstrained. To the extent that the foil constrains the comparison process, the present work adds to an understanding of similarity as "systematically fixed in context" (Medin et al., 1993, p. 272). Drawing on this notion of similarity, as well as on works on categorization and representation, I suggest that a similarity-based theory of concepts can appeal to the notion of conceptual taxonomy, in which concepts are embedded within a nested hierarchy of contrast sets, as a structure that systematically fixes the foil. Thus, such a theory is equipped to show that conceptual taxonomy constrains the similarity judgements that are intrinsically involved in categorization and facilitates the relative stability of these judgements. The proposed notion of similarity, therefore, helps alleviate the worry that the flexible nature of similarity renders it – and therefore, similarity-based theory of concepts – nonexplanatory.

4.2 Application to "causal concepts"

I have argued, above, that the notion of similarity – as construed here – provides a similarity-based theory of concepts with constraints imposed by the dynamic process of comparison of all items in the set. I conclude this article by reconsidering the worry discussed in section 1, that attempts to spell out constraints on similarity might come at the cost of restricting similarity to the perceptual level. I have already pointed to some evidence, in sections 2 and 3, showing that alignment of attributes – the process that generates the constraints at the centre of my account – need not be restricted to perceptual dimensions. But more can be said to establish that the employment of the proposed notion of similarity would enable a similarity-based theory of concepts to account for the formation and application of concepts that are quite removed from the perceptual level, such as concepts that draw on abstract, or causal knowledge.

Various studies show that alignment is not restricted to perceptual dimensions. For example, Markman and Gentner (1993b) asked participants to perform a one-shot mapping task between scenes. In this task, participants are asked to select an object in one scene that "goes with" a specific object in another scene (Markman and Gentner, 1993b, p. 435). To examine whether similarity judgements involve structural alignment, some participants were asked to rate the stimuli for similarity prior to the mapping task. When asked to align perceptual stimuli, prior performance of a similarity rating task led to an increase in subjects' reliance on relational structures rather than on local similarity during the mapping task.[40] Importantly, parallel results were obtained when participants were asked to map elements in "causal scenes". In this experiment, prior performance of a similarity rating task led to an increase in subjects' reliance on causal relational structures during the mapping task rather than on perceptual similarities, even when the perceptual similarities were salient. These results indicate that similarity comparisons involve a process of structural alignment, which can draw on relational commonalities over local perceptual similarities, and that "structural alignment provides a plausible account for similarity comparisons of both causal and perceptual relations" (Markman and Gentner, 1993b, p. 452).

The process of similarity judgement is construed, accordingly, as a specific, dynamic process in which the attributes of particulars are aligned along relational dimensions – including abstract and causal dimensions – and a search for commonalities along these dimensions is executed. The process can therefore impose constraints not only on concepts that are formed on the basis of commonalities and differences at the perceptual level, but also on concepts that build on abstract, or causal, commonalities and differences (see also discussion in Medin et al., 1993). Indeed, Gentner and Medina emphasize that it is this specific comparison process, which involves alignment, that enables us to see that different concrete situations exhibit the same relational pattern, by facilitating the abstraction of relations from the rich context of each of these situations (Gentner and Medina, 1998, p. 275). Moreover, they argue that even when we have already arrived at abstract knowledge, alignment processes are still required in order to apply this knowledge to new instances (Gentner and Medina, 1998, p. 287). Accordingly, the present notion of similarity can deliver constraints on categorization while allowing the theory to account for the formation and application of concepts at all levels of abstraction.[41]

5. Conclusion

In this article, I argued that an account of similarity that takes the foil as a relatum in similarity judgements has explanatory, conceptual and methodological advantages over an account of similarity as dyadic. First, it enables us to harness the resources of the alignment-based approach to similarity to account for foil effects, facilitating the integration of models that focus on the *process* of similarity judgements with data on the effects of the foil on the *outcome* of such judgements. Moreover, it delivers more internal constraints on similarity than the dyadic account does, without reducing the scope of similarity to the realm of the perceptual. Finally, it motivates empirical exploration and systematization of foil effects.

I then proposed that this notion of similarity is advantageous for a similarity-based account of concepts, as it brings in additional constraints on categorization. On such an account, the foil, whose variability is in large part responsible for the flexible nature of similarity judgements, is turned within the framework of conceptual taxonomy to the very element that constrains and stabilizes similarity judgements in categorization. Accordingly, a similarity-based theory of concepts can show that the process of comparison places internal constraints on similarity judgements – and therefore on categorization – thereby mitigating the worry that similarity is too flexible to be of explanatory value in a theory of concepts.

Acknowledgements

I thank Anthony Peressini and Yoon Choi for extensive and insightful comments on multiple versions of this manuscript. I also thank Fernando Martínez Manrique, M. J. García-Encinas, Neftalí Villanueva Fernández, Manolo Pinedo, Clark Wolf, Rachel Goodman, Jessica Wolfendale and Elizabeth Silverstein, as well as two anonymous reviewers, for comments and suggestions that helped improve this manuscript. Some of this material was presented to audiences at the University of Granada, Northern Illinois University, the 2016 meeting of the European Society for Philosophy and Psychology in St.

Andrews, the 2016 meeting of the Society for Philosophy and Psychology in Austin and the 2018 Context, Cognition and Communication conference in Warsaw. I thank those in attendance for their critical feedback.

Footnotes

1 I use the term "dimension" to refer to variables (e.g., 'weight'), and the terms "attribute" or "feature" to refer to a specific value (or a range of values) along a dimension (e.g., '5 kg').

2 I use single quotation marks for dimensions and attributes and small capital letters for concepts.

3 This is not to imply that similarity processing involves merely the selection and weighing of features from a pre-existing pool (for discussion, see Medin et al., [70]).

4 The term "construct" may lead to ambiguity, as it is sometimes used as a theoretical concept and sometimes as referring to a phenomenon (see discussion in Slaney and Garcia, [94]). Models of cognition that explicitly understand similarity as a construct often assume that people actually compute similarity, either directly or indirectly (see discussion in Medin et al., [70], p. 255). My own use of the term "construct" reflects the same assumption.

5 Indeed, some objections to similarity-based theories of concepts seem to hinge on the assumption that similarity is calculated over purely perceptual features (e.g., Keil, [44]).

6 Vicente and Martínez Manrique ([100]), however, point out that not all hybrid accounts are committed to the idea that the components of concepts are "semi-separable".

7 The above is not meant to provide an exhaustive list of theories of concepts. Theories not discussed here include, for example, embodied theories, which hold that concepts are encoded perceptually (e.g., Barsalou, [7], [8]; Lakoff and Johnson, [51]; Prinz, [84]; Barsalou et al., [9]; Gallese and Lakoff, [25]). Importantly, I distinguish between questions about the content of concepts and questions about their vehicle (Machery, [57]; Weiskopf, [102]; Mahon, [61]). My focus in this article is on the content of concepts, and I shall remain silent with respect to the modal-amodal debate.

8 While fully addressing the selection problem is not the goal of this article, I see the current project as contributing to that larger effort by discussing some of the constraints on similarity (for further discussion of the extent to which current similarity-based approaches are vulnerable to Goodman's worry about the flexibility of similarity, see Decock and Douven, [22]).

9 I follow Rheins's ([86]) terminology.

10 In this article, I adopt both a geometric framework and a structural alignment framework. My view is that these two approaches are quite complementary: the latter is focused on the process of similarity comparison, while the former is concerned primarily with the output of that process (see also n. 19). I note also that the general approach to similarity in the conversations I am engaging here is a naturalized one in the minimal sense of understanding the target concept of similarity as being constrained by or at least not fully independent of (in a Platonic sort of way) the process by which similarity judgements are actually made.

11 A couple of points should be made about dimensions. First, the geometric model does not require that the dimensions be continuous; it may accommodate discrete qualitative features through hierarchical featural clusters, integer values, etc. (see discussion in Gärdenfors, [26], pp. 7–8; Hahn and Chater, [39], p. 15; Krumhansl, [50], p. 460). Second, in my view, dimensions do not *have to be perceptual, and can even express high-order relations, although I should note that this goes beyond Gärdenfors's ([26]) explicit position.*

12 Gärdenfors ([26], pp. 8–9) distinguishes between phenomenal (or psychological) dimensions and scientific (or theoretical) dimensions; the scales of the phenomenal dimensions can stretch or shrink.

13 I borrow the term "diagnosticity" from Tversky and Gati ([99]).

14 The figure is used only to illustrate the effects of local density on similarity judgements as proposed by Krumhansl (1978). For present purposes, I ignore other factors that might affect one's similarity estimate for 'x' and 'y', such as the spatial arrangement of the objects, etc.

15 Let us suppose, for the sake of the example, that 'cost of living' designates a single dimension.

16 It is not my intention here to argue for any specific position with respect to gradable adjectives. Rather, I find one particular line of research useful in illustrating my own argument about similarity.

17 Klein (1980, p. 13) defines a comparison class as "a subset of the universe of discourse which is picked out relative to a context of use".

18 In general, context factors that are excluded from the model are non-essential and/or non-systematizable factors that have to be fixed (specified) before a systematic relationship can be formulated. Thus, when some of the factors that have heretofore been relegated to "the non-systematizable external context" become understood as ubiquitous and systematizable, then they are usefully incorporated into the model.

19 My working assumption here is that insight from alignment-based models could complement geometric accounts, as it could explain the process by which representations of particulars are placed along different dimensions in conceptual space for similarity judgements. This is in line with Goldstone et al.'s (1997, p. 251) suggestion that alignment-based models and geometric (or featural) models may be integrated, as the former provide the processing aspects of similarity judgements while the latter provide the structural aspects of the resultant similarity.

20 This is not to say that the meaning of the statement "'x' and 'y' are similar" amounts to "'x' and 'y' are closer to each other than they are to 'f'." My project here is not a semantic one, and I make no claims about the meaning of similarity statements. Rather, my focus is on the objects that are compared, and the relationship – among those objects – that has to be grasped in order to arrive at similarity judgements. I thank Neftalí Villanueva Fernández and Manolo Pinedo for pressing me on this point.

21 In what follows, I continue to refer to the foil as an element of the broad umbrella of "context", while also maintaining that the foil is one of the relata in the similarity relationship. I employ the term in this way (i) for ease of relating the present view to other works, and (ii) to acknowledge that the foil is not the explicit target of similarity judgements, and hence does, in this sense, function in the background much like context does.

22 It should be clear, at this point, that the foil need not be a single object. This might lead to a worry about whether the notion of similarity proposed here is truly "triadic". Note, however, that the same issue is faced by a "dyadic" view of similarity as well: the dyadic notion is often applied to a group of objects (e.g., in the case of categories), and some mechanism of averaging is assumed. The triadic view can help itself to this assumption, as well. My concern here, however, is with the most minimal relation required to produce a similarity judgement.

23 My suggestion about the effect of the foil on the alignment of the target is supported indirectly by the data discussed above; obtaining direct evidence would require testing the effects of manipulations of the foil on the alignment of targets (via, e.g., a mapping task).

24This parallels a point made by Yee and Thompson-Schill ([103]) about the role of context in conceptualization.

25 Note that in this example, as in (i), the role of the foil in specifying the relevant dimension is bypassed by an explicit statement of this dimension. I return to this issue below.

26 Of course, the reliance on an implicit foil may lead to ambiguities in similarity judgements (just as the reliance on implicit comparison classes may lead to ambiguities in the application of relative gradable adjectives) as the proper foil may not always be clear from the general context. In section 3, I suggest that in the context of conceptual hierarchy, the taxonomic structure of a conceptual system helps reduce such ambiguity.

27 This again parallels discussions about the role of comparison classes in the representation of gradable adjectives. Graff ([37]) points out that the judgement "John is tall for a basketball player" does not depend merely on John's height in relation to the average height of basketball players. Rather, it depends on what we know about the typical height of basketball players. Here, too, we need to know something about the distribution of values, along the relevant dimension, within the comparison class. Graff's position echoes well-known studies in psychology. Parducci ([81]), for example, shows that subjects' frame of reference for judgements such as "small" and "large" does not depend merely on the average value of the stimuli set, but on both the extreme values in the set and the frequency distribution of these values.

28 My use of conceptual taxonomy is not meant as a statement about the organization of conceptual knowledge in the brain (for discussion, see Mahon and Caramazza, [62]). Rather, the taxonomy merely describes the way that classifiers take categories of things in the world to be related to each other.

29 Treating all contrast classes within the same inclusive category as equally important in similarity judgement is an oversimplification for the sake of brevity. Different categories within the contrast set may differentially affect similarity judgements with respect to a given category. While psychologists often define contrast classes as the "other categories from the same parent category" (Markman and Wisniewski, [67], p. 57, n. 1), the actual methods they use to operationalize this notion often generate only those categories, within the contrast set, that are the closest to the target category (see, e.g., Rosch and Mervis, [91]; Malt and Johnson, [63]; Markman and Wisniewski, [67]). For present purposes, I put these details aside.

30 This is not to imply that all items are compared simultaneously (see discussion in Clapper, [18]).

31 Rosch et al. showed that, for concrete objects, taxonomies are "structured such that there is generally one level of abstraction at which the most basic category cuts can be made." They proposed that the basic level of inclusiveness in a taxonomy is "the level at which categories carry the most information, possess the highest cue validity, and are thus the most differentiated from one another" (Rosch et al., [92], p. 383).

32 Ease of category learning is taken as indicative of the degree of judged similarity. The assumption here is that it is easy to learn categories when the similarity among category members is high and the similarity among members of different categories is low. Conversely, it is difficult to *learn categories when the similarity among category members is low and the similarity among members of different categories is high.*

33 For consistency, I use "commonalities" where Lassaline and Murphy used "matches". As an aside, it is worth noting that a description of features as "commonalities" or as "matching" is helpful when we wish to focus on questions about the selection and weighing of dimensions in similarity judgements, while putting aside calculations internal to each dimension. We should keep in mind, however, that this is an oversimplification; outside the lab setting, people construct categories that are likely to employ dimensions with non-binary values (see discussion in Andrews et al., [4]).

34 The term "diagnostic" is used by Tversky ([98]). Other terms used to express this notion are "distinguishing" features (Cree and McRae, [21]), "distinctiveness" (Garrard et al., [27]; Cree et al., [20]), "informativeness" (Devlin et al., [23]) and "cue validity" (Bourne Jr and Restle, [15]; Rosch and Mervis, [91]). These authors, however, differ in their view of diagnosticity (or distinctiveness) of features as a local or a global variable. Some focus on the global distinctiveness of features. For example, the feature 'moos' is globally distinctive for members of cow, as it distinguishes them from members of all other basic level categories (Cree et al., [20]). Others focus on the distinctiveness of features in relation to a specific contrast set. For example, while the feature 'has wings' is not distinctive within the broad category of birds, since it is shared by members of all of its subordinate categories (e.g., robin, eagle), it is distinctive within the broad category of vehicles (Garrard et al., [27]; see also discussion of cue validity in Rosch and Mervis, [91]). In this article, my focus is on locally distinctive features.

35 This is not to imply that a taxonomy is formed independently of any prior conceptualization; previously formed concepts may be used in the formation of new contrast sets. (I thank an anonymous reviewer for bringing up this point.) I shall only mention a few examples of such possible dependencies. First, previously formed concepts may be used in the process of alignment or the identification of similarities and differences when a new taxonomy is formed. For example, Billman and Davies ([10]) asked subjects to form novel categories for sorting images of aliens. Aliens in the images presented to participants differed from each other with respect to the shape of their tail, the shape of their nose, etc. Performance in the sorting task may have been facilitated by subjects' familiarity with concepts such as tail, nose, etc. Second, the order of concept formation may depend on the distance from the perceptual level. For example, Newen and Bartels ([77]) distinguish between primitive perception-based concepts (e.g., red) and more abstract concepts (e.g., colour), arguing that abstract concepts are introduced through representations that are based on more primitive concepts. Third, children learn basic-level categories (i.e., categories at an intermediate level of inclusiveness), such as chair, before learning categories at the superordinate and subordinate levels, such as furniture and kitchenchair, respectively (Rosch et al., [92]; Mervis and Crisafi, [75]). Note that these different ways in which the formation of new categories may depend on prior taxonomy reflect a hierarchy in category learning, which does not necessarily lead to circularity. For example, we may need to have the concept red (or some other specific colour concepts) before forming the concept colour, but we do not need to have the concept colour before forming the concept red (see discussion in Newen and Bartels, [77]).

36 The set may include the items directly under consideration, but also items retrieved or constructed based on prior experience (see discussion in Medin et al., [71]).

37 See also Mellers and Biagini ([74]), who propose that when items receive close values along one dimension, the differences between them along another dimension are enhanced.

38 Note that Rosch here talks about the perceived world, rather than "a metaphysical world without a knower" (Rosch, 1978, p. 29). The assumption here is that what we perceive as "information rich bundles" is constrained by our perceptual system; we are hard-wired to notice some dimensions over others.

39 I do not mean to imply that concepts are rigid constructs that do not change over time (indeed, I take concepts as fundamentally dynamic; see Bloch-Mullins, forthcoming). Gärdenfors ([26], ch. 4) provides a useful discussion of the way in which observation of new items, and their incorporation into the taxonomy, may affect subsequent similarity judgements.

40 Similarity comparisons generally led to an increase in relational alignment in both sparse and rich stimuli. However, the preference for relational alignment over local object feature match was reversed when very rich local similarities were placed in opposition to a simple relational structure (Markman and Gentner, [65]).

41 This is not to deny that there may be important differences between concepts that draw on perceptual features and concepts that build on abstract or causal knowledge (e.g., Newen and Bartels, [77]; Dove, [24]). Moreover, the specific way by which alignment constrains categorization may vary at different levels of a taxonomic hierarchy (e.g., Markman and Wisniewski, [67]).

References

Ahn, W. K., Kim, N. S., Lassaline, M. E., and Dennis, M. J. (2000) " Causal Status as a Determinant of Feature Centrality." Cognitive Psychology 41 : 361 – 416.

Ahn, W. K., Marsh, J. K., Luhmann, C. C., and Lee, K. (2002) " Effect of Theory-Based Feature Correlations on Typicality Judgments." Memory and Cognition 30 (1): 107 – 118.

Anderson, J. R. and Betz, J. (2001) " A Hybrid Model of Categorization." Psychonomic Bulletin and Review 8 (4): 629 – 647.

Andrews, J. K., Livingston, K. R., and Kurtz, K. J. (2011) " Category Learning in the Context of Co-Presented Items." Cognitive Processing 12 (2): 161 – 175.

Barsalou, L. W. (1982) " Context-Independent and Context-Dependent Information in Concepts." Memory and Cognition 10 (1): 82 – 93.

Barsalou, L. W. (1993) " Flexibility, Structure, and Linguistic Vagary in Concepts: Manifestations of a Compositional System of Perceptual Symbols." In A. F. Collins, S. E. Gathercole, M. A. Conway and P. E. Morris (eds), Theories of Memory, pp. 29 – 101. London : Lawrence Erlbaum Associates.

Barsalou, L. W. (1999) " Perceptual Symbol Systems." Behavioral and Brain Sciences 22 : 577 – 660.

Barsalou, L. W. (2008) " Grounded Cognition." Annual Review of Psychology 59 : 617 – 645.

Barsalou, L. W., Simmons, K. W., Barbey, A. K., and Wilson, C. D. (2003) " Grounding Conceptual Knowledge in Modality-Specific Systems." Trends in Cognitive Sciences 7 (2): 84 – 91.

Billman, D. and Davies, J. (2005) " Consistent Contrast and Correlation in Free Sorting." American Journal of Psychology 118 (3): 353 – 383.

Billman, D. and Dávila, D. (2001) " Consistent Contrast Aids Concept Learning." Memory and Cognition 29 (7): 1022 – 1035.

Bloch-Mullins, C. L. (2015) " Foundational Questions about Concepts: Context-Sensitivity and Embodiment." Philosophy Compass 10 : 940 – 952.

Bloch-Mullins, C. L. (2018) " Bridging the Gap between Similarity and Causality: An Integrated Approach to Concepts." British Journal for the Philosophy of Science 69 (3): 605 – 632.

Bloch-Mullins, C. L. (forthcoming). "Scientific Concepts as Forward-Looking: How Taxonomic Structure Facilitates Conceptual Development." Journal of the Philosophy of History.

Bourne, L. E. Jr. and Restle, F. (1959) " Mathematical Theory of Concept Identification." Psychological Review 66 (5): 278 – 296.

Carey, S. E. (1985) Conceptual Change in Childhood. Cambridge, MA : MIT Press.

Casasanto, D. and Lupyan, H. (2015) " All Concepts Are Ad Hoc Concepts." In E. Margolis and S. Laurence (eds), The Conceptual Mind: New Directions in the Study of Concepts, pp. 543 – 566. Cambridge, MA : MIT Press.

Clapper, J. P. (2017) " Alignability-Based Free Categorization." Cognition 162 : 87 – 102.

Corter, J. E. (1987) " Similarity, Confusability, and the Density Hypothesis." Journal of Experimental Psychology: General 116 (3): 238 – 249.

Cree, G. S., McNorgan, C., and McRae, K. (2006) " Distinctive Features Hold a Privileged Status in the Computation of Word Meaning: Implications for Theories of Semantic Memory." Journal of Experimental Psychology: Learning, Memory, and Cognition 32 (4): 643 – 658.

Cree, G. S. and McRae, K. (2003) " Analyzing the Factors Underlying the Structure and Computation of the Meaning of Chipmunk, Cherry, Chisel, Cheese, and Cello (and Many Other Such Concrete Nouns)." Journal of Experimental Psychology: General 132 (2): 163 – 201.

Decock, L. and Douven, I. (2011) " Similarity after Goodman." Review of Philosophy and Psychology 2 (1): 61 – 75.

Devlin, J. T., Gonnerman, L. M., Anderson, E. S., and Seidenberg, M. S. (1998) " Category-Specific Semantic Deficits in Focal and Widespread Damage: A Computational Account." Journal of Cognitive Neuroscience 10 (1): 77 – 94.

Dove, G. (2009) " Beyond Perceptual Symbols: A Call for Representational Pluralism." Cognition 110 : 412 – 431.

Gallese, V. and Lakoff, G. (2005) " The Brain's Concepts: The Role of the Sensory-Motor System in Conceptual Knowledge." Cognitive Neuropsychology 22 (3/4): 455 – 479.

Gärdenfors, P. (2000) Conceptual Spaces: The Geometry of Thought. Cambridge, MA : MIT Press.

Garrard, P., Ralph, M. A. L., Hodges, J. R., and Patterson, K. (2001) " Prototypicality, Distinctiveness, and Intercorrelation: Analyses of the Semantic Attributes of Living and Nonliving Concepts." Cognitive Neuropsychology 18 (2): 125 – 174.

Gelman, S. A. and Markman, E. M. (1986) " Categories and Induction in Young Children." Cognition 23 : 183 – 209.

Gentner, D. and Gunn, V. (2001) " Structural Alignment Facilitates the Noticing of Differences." Memory and Cognition 29 (4): 565 – 577.

Gentner, D. and Markman, A. B. (1994) " Structural Alignment in Comparison: No Difference without Similarity." Psychological Science 5 (3): 152 – 158.

Gentner, D. and Medina, J. (1998) " Similarity and the Development of Rules." Cognition 65 (2–3): 263 – 297.

Gentner, D. and Rattermann, M. J. (1991) " Language and the Career of Similarity." In S. A. Gelman and J. P. Byrnes (eds), Perspectives on Language and Thought: Interrelations in Development, pp. 225 – 277. London : Cambridge University Press.

Goldstone, R. L. (1994a) " Similarity, Interactive Activation, and Mapping." Journal of Experimental Psychology: Learning, Memory, and Cognition 20 (1): 3 – 28.

Goldstone, R. L. (1994b) " The Role of Similarity in Categorization: Providing a Groundwork." Cognition 52 : 125 – 157.

Goldstone, R. L., Medin, D. L., and Halberstadt, J. (1997) " Similarity in Context." Memory and Cognition 25 : 237 – 255.

Goodman, N. (1972) " Seven Strictures on Similarity." In N. Goodman (ed.), Problems and Projects, pp. 437 – 447. Indianapolis, IN : Bobbs-Merrill.

Graff, D. (2000) " Shifting Sands: An Interest-Relative Theory of Vagueness." Philosophical Topics 28 (1): 45 – 81.

Griffiths, P. E. (1997) What Emotions Really Are: The Problem of Psychological Categories. Chicago, IL : University of Chicago Press.

Hahn, U. and Chater, N. (1997) " Concepts and Similarity." In K. Lamberts and D. Shanks (eds), Knowledge, Concepts and Categories, pp. 43 – 92. New York : Psychology Press.

Hampton, J. A. (1998) " Similarity-Based Categorization and Fuzziness of Natural Categories." Cognition 65 : 137 – 165.

Hsu, N. S., Schlichting, M. L., and Thompson-Schill, S. L. (2014) " Feature Diagnosticity Affects Representations of Novel and Familiar Objects." Journal of Cognitive Neuroscience 26 (12): 2735 – 2749.

Hull, D. L. (1988) Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science. Chicago, IL : University of Chicago Press.

Kay, P. (1971) " Taxonomy and Semantic Contrast." Language 47 (4): 866 – 887.

Keil, F. C. (1986) " The Acquisition of Natural Kind and Artifact Terms." In W. Dimopoulos and A. Marras (eds), Language Learning and Concept Acquisition: Foundational Issues, pp. 133 – 153. Norwood, NJ : Ablex.

Keil, F. C. (1989) Concepts, Kinds, and Cognitive Development. Cambridge, MA : MIT Press.

Keil, F. C. (1994) " Explanation, Association, and the Acquisition of Word Meaning." Lingua 92 : 169 – 196.

Kennedy, C. (2007) " Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives." Linguistics and Philosophy 30 (1): 1 – 45.

King, D. L. and Atef-Vahid, M.-K. (1986) " Two Extensions of the Anchor-Range Effect." Perception & Psychophysics 39 (2): 96 – 104.

Klein, E. (1980) " A Semantics for Positive and Comparative Adjectives." Linguistics and Philosophy 4 (1): 1 – 45.

Krumhansl, C. L. (1978) " Concerning the Applicability of Geometric Models to Similarity Data: The Interrelationship between Similarity and Spatial Density." Psychological Review 85 (5): 445 – 463.

Lakoff, G. and Johnson, M. (1999) Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought. New York : Basic Books.
Lassaline, M. E. and Murphy, G. L. (1998) " Alignment and Category Learning." Journal of Experimental Psychology: Learning, Memory, and Cognition 24 (1): 144 – 160.
Laurence, S. and Margolis, E. (1999) " Concepts and Cognitive Science." In E. Margolis and S. Laurence (eds), Concepts: Core Readings, pp. 3 – 81. Cambridge, MA : MIT Press.
Lebois, L. A. M., Wilson-Mendenhall, C. D., and Barsalou, L. W. (2015) " Are Automatic Conceptual Cores the Gold Standard of Semantic Processing? The Context-Dependence of Spatial Meaning in Grounded Congruency Effects." Cognitive Science 39 : 1764 – 1801.
Löhr, G. (2017) " Abstract Concepts, Compositionality, and the Contextualism-Invariantism Debate." Philosophical Psychology 30 (6): 689 – 710.
Machery, E. (2005) " Concepts Are Not a Natural Kind." Philosophy of Science 72 (3): 444 – 467.
Machery, E. (2007) " Concept Empiricism: A Methodological Critique." Cognition 104 : 19 – 46.
Machery, E. (2009) Doing without Concepts. New York : Oxford University Press.
Machery, E. (2014) " Concepts: Investigating the Heterogeneity Hypothesis." In J. Sytsma (ed.), Advances in Experimental Philosophy of Mind, pp. 203 – 221. London : Bloomsbury Academic.
Machery, E. (2015) " By Default: Concepts Are Accessed in a Context-Independent Manner." In E. Margolis and S. Laurence (eds), The Conceptual Mind: New Directions in the Study of Concepts, pp. 567 – 588. Cambridge, MA : MIT Press.
Mahon, B. Z. (2015) " What Is Embodied about Cognition? " Language, Cognition and Neuroscience 30 (4): 420 – 429.
Mahon, B. Z. and Caramazza, A. (2009) " Concepts and Categories: A Cognitive Neuropsychological Perspective." Annual Review of Psychology 60 : 27 – 51.
Malt, B. C. and Johnson, E. C. (1992) " Do Artifact Concepts Have Cores? " Journal of Memory and Language 31 (2): 195 – 217.
Markman, A. B. and Gentner, D. (1993a) " Splitting the Differences: A Structural Alignment View of Similarity." Journal of Memory and Language 32 (4): 517 – 535.
Markman, A. B. and Gentner, D. (1993b) " Structural Alignment during Similarity Comparisons." Cognitive Psychology 25 (4): 431 – 467.
Markman, A. B. and Gentner, D. (1996) " Commonalities and Differences in Similarity Comparisons." Memory and Cognition 24 (2): 235 – 249.
Markman, A. B. and Wisniewski, E. J. (1997) " Similar and Different: The Differentiation of Basic-Level Categories." Journal of Experimental Psychology: Learning, Memory, and Cognition 23 (1): 54 – 70.
Mazzone, M. and Lalumera, E. (2010) " Concepts: Stored or Created? " Minds and Machines 20 (1): 47 – 68.
Medin, D. L. (1989) " Concepts and Conceptual Structure." American Psychologist 44 (12): 1469 – 1481.
Medin, D. L., Goldstone, R. L., and Gentner, D. (1993) " Respects for Similarity." Psychological Review 100 (2): 254 – 278.
Medin, D. L., Goldstone, R. L., and Markman, A. B. (1995) " Comparison and Choice: Relations between Similarity Processes and Decision Processes." Psychonomic Bulletin and Review 2 (1): 1 – 19.
Medin, D. L. and Schaffer, M. M. (1978) " Context Theory of Classification Learning." Psychological Review 85 (3): 207 – 238.

Medin, D. L., Wattenmaker, W. D., and Hampson, S. E. (1987) " Family Resemblance, Conceptual Cohesiveness, and Category Construction." Cognitive Psychology 19 (2): 242 – 279.

Mellers, B. A. and Biagini, K. (1994) " Similarity and Choice." Psychological Review 101 (3): 505 – 518.

Mervis, C. B. and Crisafi, M. A. (1982) " Order of Acquisition of Subordinate-, Basic-, and Superordinate-Level Categories." Child Development 53 (1): 258 – 266.

Murphy, G. L. and Medin, D. L. (1985) " The Role of Theories in Conceptual Coherence." Psychological Review 92 (3): 289 – 316.

Newen, A. and Bartels, A. (2007) " Animal Minds and the Possession of Concepts." Philosophical Psychology 20 (3): 283 – 308.

Nosofsky, R. M. (1986) " Attention, Similarity, and the Identification-Categorization Relationship." Journal of Experimental Psychology: General 115 (1): 39 – 57.

Nosofsky, R. M. (1988) " Exemplar-Based Accounts of Relations between Classification, Recognition, and Typicality." Journal of Experimental Psychology: Learning, Memory, and Cognition 14 (4): 700 – 708.

Osherson, D. N. and Smith, E. E. (1981) " On the Adequacy of Prototype Theory as a Theory of Concepts." Cognition 9 : 35 – 58.

Parducci, A. (1965) " Category Judgment: A Range-Frequency Model." Psychological Review 72 (6): 407 – 418.

Piccinini, G. and Scott, S. (2006) " Splitting Concepts." Philosophy of Science 73 (4): 390 – 409.

Posner, M. I. and Keele, S. W. (1968) " On the Genesis of Abstract Ideas." Journal of Experimental Psychology 77 (3): 353 – 363.

Prinz, J. J. (2002) Furnishing the Mind: Concepts and Their Perceptual Basis. Cambridge, MA : MIT Press.

Quine, W. V. (1969) " Natural Kinds." Ontological Relativity and Other Essays, pp. 114 – 138. New York : Columbia University Press.

Rheins, J. G. (2011) " Similarity and Species Concepts." In J. K. Campbell, M. O'Rourke and M. H. Slater (eds), Carving Nature at Its Joints: Natural Kinds in Metaphysics and Science, pp. 253 – 288. Cambridge, MA : MIT Press.

Rice, C. (2016) " Concepts as Pluralistic Hybrids." Philosophy and Phenomenological Research 92 (3): 597 – 619.

Rips, L. J. (1989) " Similarity, Typicality, and Categorization." In S. Vosniadou and A. Ortony (eds), Similarity and Analogical Reasoning, pp. 21 – 59. Cambridge, MA : Cambridge University Press.

Rips, L. J., Shoben, E. J., and Smith, E. E. (1973) " Semantic Distance and the Verification of Semantic Relations." Journal of Verbal Learning and Verbal Behavior 12 (1): 1 – 20.

Rosch, E. (1978) " Principles of Categorization." In E. Rosch and B. B. Lloyd (eds), Cognition and Categorization, pp. 27 – 48. Hillsdale, NJ : Lawrence Erlbaum.

Rosch, E. and Mervis, C. B. (1975) " Family Resemblances: Studies in the Internal Structure of Categories." Cognitive Psychology 7 : 573 – 605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., and Boyes-Braem, P. (1976a) " Basic Objects in Natural Categories." Cognitive Psychology 8 (3): 382 – 439.

Rosch, E., Simpson, C., and Miller, R. S. (1976b) " Structural Bases of Typicality Effects." Journal of Experimental Psychology: Human Perception and Performance 2 (4): 491 – 502.

Slaney, K. L. and Garcia, D. A. (2015) " Constructing Psychological Objects: The Rhetoric of Constructs." Journal of Theoretical and Philosophical Psychology 35 (4): 244 – 259.
Sloman, S. A. and Rips, L. J. (1998) " Similarity as an Explanatory Construct." Cognition 65 : 87 – 101.
Smith, E. E., Shoben, E. J., and Rips, L. J. (1974) " Structure and Process in Semantic Memory: A Featural Model for Semantic Decisions." Psychological Review 81 (3): 214 – 241.
Torgerson, W. S. (1965) " Multidimensional Scaling of Similarity." Psychometrika 30 (4): 379 – 393.
Tversky, A. (1977) " Features of Similarity." Psychological Reviews 84 (4): 327 – 352.
Tversky, A. and Gati, I. (1978) " Studies of Similarity." In E. Rosch and B. L. Lloyd (eds), Cognition and Categorization, pp. 79 – 98. Hillsdale, NJ : Lawrence Erlbaum.
Vicente, A. and Martínez Manrique, F. (2016) The Big Concepts Paper: A Defence of Hybridism. The British Journal for the Philosophy of Science 67 (1): 59 – 88.
Weiskopf, D. A. (2009) " The Plurality of Concepts." Synthese 169 : 145 – 173.
Weiskopf, D. A. (2010) " Embodied Cognition and Linguistic Comprehension." Studies in History and Philosophy of Science Part A 41 (3): 294 – 304.
Yee, E. and Thompson-Schill, S. L. (2016) " Putting Concepts into Context." Psychonomic Bulletin and Review 23 (4): 1015 – 1027.