What makes for conceptual similarity?

Article *in* Theoria Beograd · January 2017

DOI: 10.2298/THEO1704077S

Olivera Savic
The Ohio State University

THEORIA 4
BIBLID 0369–2485 : (2017) : 60: str. 77-90

https://doi.org/10.2298/THEO1704077S
Originalni naučni rad
Original Scientific Paper

Olivera Savić and Hope Sample

WHAT MAKES FOR CONCEPTUAL SIMILARITY?

ABSTRACT: A significant number of phenomena in psychology is explained in terms of similarity. While the term has found to be useful in understanding and defining other phenomena, the similarity itself remains to be poorly understood and defined. Here we aim to discuss the current status of the concept of similarity as it is applied to research in cognitive psychology and cognitive neuroscience.

KEYWORDS: conceptual similarity; structural alignment model; latent semantic analyses

1. Introduction

As it was accurately noted by Quine (1969; p. 116) *"For surely there is nothing more basic to thought and language than our sense of similarity; our sorting of things into kinds."* Sense of similarity is crucial for human cognition. It is a foundation of cognitive processes that range from early perceptual acts to complex decision making. We are able to recognize a novel item as an apple, to make inferences about its non-observable characteristics (edible, healthy, sweet), make decisions on how to interact with it and predict consequences of our actions towards novel, never seen before objects. We are able to do so based on our sense of similarity between the novel item and items that we are familiar with, though there is much controversy over how to understand this sense. Being able to organize concepts by their similarity is a fundamental operation in the human mind and thus it is crucial for cognitive psychology to understand the origin of the human sense of similarity.

In this paper, we aim to review the current status of conceptual similarity as it is used in cognitive psychology. Cognitive psychology is at its theoretical crossroads with respect to its approach to conceptual similarity, which makes this discussion timely. Historically assessments of conceptual similarity were seen as a comparison process between the intrinsic features of the objects (Tversky, 1977; Gentner, 1983). In recent years there has been a shift from an intrinsic features view of conceptual similarity to a view that could account for relational information that is extrinsic to the compared items (Wisniewski & Bassok, 1999; Jones & Love, 2007).

As a "very keel and backbone of our thinking" (William James 1890, p. 459), it is reasonable to expect that our sense of similarity should be invariant and stable in order to reliably govern our cognitive system. However, this expectation has proven to be false by a variety of empirical results showing that our sense of similarity is extremely flexible and context-dependent (Estes, Golonka, & Jones, 2011; Medin, Goldstone, & Gentner, 1993; Medin & Rips, 2005). These studies suggest that whether it will make sense to compare apples and oranges and perceive them as similar may depend on our current goals, surrounding context, cultural, and developmental factors. In addition to these issues, problems in defining factors that affect similarity also arise from use of the same construct – similarity - to describe different types of closeness between concepts. We effortlessly perceive a letter and an email as similar because they are both used to transfer a message to someone else. At the same time, a letter and a diary could also be perceived as similar, but in this case, based on the overlap in their perceptual characteristics (e.g. made of paper, can be written in, and so on). Although we use the same term – similarity - to describe relationship between both pairs of concepts, some authors suggest that these similarities may not be the same thing and that they may rely on different neural mechanisms (Chen et al, 2013).

While, invariance of similarity made some authors question its usefulness as an explanatory tool, others were more optimistic and argued that findings of variability in similarity judgments disguised and drew attention of researchers from systematic regularities of the process (e.g. Bassok & Madin, 1997). On this view, by studying of systematic constraints, by better understanding of selection and weighting of the relevant properties we could explain variability in similarity judgments and understand what makes different similarities similar. One of the main reasons variability in similarity was seen as problematic is that it contrasted the view of concepts that dominated research in cognitive psychology for decades - the view of concepts as solely determined by intrinsic features (Tulving, 1972). However, the view of the nature and neural organization of concepts is changing towards one in which concepts encode not only information about intrinsic features of entities they represent, but also information about connections with other concepts (see Murphy, 2001; Murphy, 2002; Sachs, Weis, Zellagui, Huber, Zvyagintsev, Mathiak, & Kircher, 2008; Mirman, Walker & Graziano, 2011). This fundamental change in view of nature of concepts has potential to shed new light on challenges in explaining variability in similarity assessment and it offers an opportunity for reconciliation between different faces of similarity (Goldstone, 1995). We will suggest that this new approach is also a fruitful one.

We start with a brief review of the models of similarity that illustrate with these two broad frameworks and the empirical findings that provoked development within these frameworks. Further, we discuss the implications of these different views of conceptual similarity for models and theories of conceptual knowledge organization. Finally, the review is followed by a discussion that aims to bridge the gap between intrinsic features models of similarity (Tversky, 1977; Gentner, 1983) and empirical findings of effects of extrinsic features information (Wisniewski & Bassok, 1999; Jones & Love, 2007).

2. Background on Conceptual Similarity in Cognitive Psychology

Before we go into the main discussion it is important to make some clarifications regarding our standpoint. To begin with, the issues about conceptual similarity that are emphasized in cognitive psychology should be distinguished from other sorts of issues one might raise about conceptual similarity. Cognitive psychology aims to determine similarity between two concepts in a single system, rather than determining the similarity between concepts of the same entity in two separate systems. Thus, cognitive psychology seeks to understand what makes the concept of an apple (in my head) similar to the concept of a pear (in my head, again), rather than what makes it possible for concept of apple in my head and concept of apple in your head to be similar enough so we understand each other when using these concepts. While there are some interesting overlaps in questions raised by these two faces of similarity, here we will restrict our discussion to the problem of similarity between concepts within a single system.

While we acknowledge that the question of primacy, whether (or to what extent) mental states are determined by experience or vice versa, is a matter of an open debate, for the sake of current discussion we will assume primacy of empirical concepts of entities, thus assuming that concepts are inductively created mental structures. In many theories of conceptual learning, mental representations are assumed to be completely based on experience with the entities one has been exposed to, hence (while not necessarily explicitly stated) the assumption of primacy of experience is implicitly made (see Estes, 1994; Medin & Schaffer, 1978; Nosofsky, 1984).

3. Intrinsic Features Models of Similarity

One of the most intuitive models of similarity is the one suggested by Tversky (1977; Tversky & Gati, 1978). Tversky's *contrast model* defines perceived similarity as a feature matching process. Similarity between two objects increases as the number of the common features increases and/or the number of distinctive features decreases. More precisely, Tversky's model claims that highly similar objects are those that share all the same intrinsic properties and have no unique intrinsic properties. On this model, items that do not share intrinsic properties and hence have all distinctive intrinsic properties, are on the other extreme, being completely different.

However, we know that people can perceive similarity between entities that have no overlapping intrinsic properties. For example, Rattermann and Gentner (1987) found that stories about different characters in similar roles were rated as more similar than stories about similar characters in different roles. A similar result was found when figural stimuli were compared. To take a simplified example, imagine similarity between '▼●▼' and '●▼▼' is compared. Tversky's contrast model treats objects as mere lists of intrinsic properties and does not take into account the intrinsic relational qualities of objects. Based on Tversky's *contrast model,* '▼●▼' and '●▼▼' are identical since they share all the same intrinsic properties and there are no distinctive intrinsic properties. Yet, human participants judge similarity between '▼●▼' and '●▼▼' (same elements, order changed) to be low and even lower than similarity between '▼●▼' and '■●■' (elements differ, order maintained). This finding proves that similarity judgments are sensitive to the relational structure of the elements of objects in the sense that similarity of structure is sometimes weighted more than overlap in intrinsic properties in human similarity judgments (Rattermann & Gentner, 1987). These findings could not be predicted by models that compare the intrinsic properties of the elements of an object in isolation from one another.

Inspired by studies of analogical reasoning, Markman and Gentner (1993) proposed the *structural alignment model* of similarity. As in Tversky's model (1977) determining the similarity requires recognizing commonalities and differences of a pair. Improving upon the *contrast model,* the *structural alignment model* takes into account not only information about intrinsic properties, but also information about the intrinsic relations between the parts of an object. This difference is fundamental. While contrast model assumes that similarity estimate can be made based on comparison of mere lists of intrinsic properties, structural alignment model assumes that what is compared are structures of intrinsic features. In other words, sense of similarity is based on comparison of interconnected elements rather than independent elements. The structural alignment model accounts for two types of intrinsic commonalties: properties and structural relations.

More specifically, the structural alignment model predicts that similarity comparison starts with alignment of the structured representations. Parallel to case of analogical mapping, this process of alignment places elements of the structured representation in correspondence with one another (Markman & Gentner, 1993). Based on this mapping, the system further identifies commonalities and differences by comparing corresponding elements. Back to the example given above, '▼●▼' and '■●■' would first be placed in correspondence based on their relational structure (e.g. order of elements). The fact that both configurations have symmetric AbA structure would be a relational commonality of the pair. Further comparison would show that central object is the circle in both configurations, which is their intrinsic property commonality. Finally, differences in intrinsic properties would be detected. The outer elements are triangles in the first and squares in the second configuration. Based on the main assumption of the structural alignment model, it is predicted that importance of feature overlap is dependent on structural relations among intrinsic properties of objects (Gentner, 1983; Markman & Gentner, 1993).

The *contrast model* (Tversky, 1977) and *structural alignment model* (Markman & Gentner, 1993) share the view of concepts as they are described in the traditional theories of conceptual knowledge organization (Quillian, 1968; Collins & Loftus, 1975; Rosch, 1975). Concepts capture the key intrinsic properties of entities and are further organized similarly as the scientific classifications in biology, where items are grouped and categorized based on their commonalities. Conceptual knowledge is seen as general knowledge that is free of context in which it is acquired (Tulving, 1972). Most importantly for our discussion, both traditional theories of conceptual knowledge organization and presented models of similarity assume that concepts are defined solely based on intrinsic properties and relations of items they represent and that extrinsic relations (relations between separate items) affect neither the principles of conceptual organization nor similarity judgments. The next section presents evidence that runs counter to these assumptions by demonstrating importance of extrinsic relational information.

4. Extrinsic Features Models of Similarity

If the meaning of concepts is determined (at least in part) by other concepts, then concepts capture information about properties and relations that are extrinsic to the items they represent. From the perspective of the traditional model of concepts (e.g. Quillian, 1968; Rosch, 1975; Rogers & McClelland, 2004), such extrinsic information is neglected, which prevented them from acknowledging the role of extrinsic features in similarity judgments. The advances in our understanding of the role of extrinsic features in assessing similarity dovetails with the transition to understanding the meaning of concepts holistically in terms of their relations to other concepts. We will review studies that illustrate different efforts to explicitly test or model the hypothesis that extrinsic information affects similarity judgments.

In their seminal paper, Landauer and Dumais (1997) presented a computational model that extracts meaning of words solely based on the contexts they appear in. To go back to the beginning of the twentieth century, quite similar idea was proposed by Ferdinand de Saussure (1915) who argued that concepts can be completely "negatively defined" in terms of other concepts. In other words, Saussure (1915) argued for idea that concepts are defined in terms of other concepts, thus by their extrinsic relations, rather than ("positively") in terms of their content that describes intrinsic features of objects. Based on a large corpus of texts and powerful mathematical analysis tools, LSA (Latent Semantic Analyses, Landauer & Dumais, 1997) represents concepts in terms of their occurrences in particular text units (e.g. sentences). Words that occur in similar contexts are thus represented similarly. Although LSA based measure of similarity captures only overlap in neighboring concepts and has no information about intrinsic properties of concepts, it has proven to successfully predict complex phenomena in cognition and language, from semantic priming effects to the typical vocabulary growth rate of school children (Landauer & Dumais, 1997).

Latent Semantic Analyses model of Landauer and Dumais (1997) stands in high contrast with Tversky's (1977) model of similarity. While Tversky's model sees concepts as represented exclusively in terms of intrinsic features of entities, LSA completely ignores the intrinsic features of concepts. What makes these two models completely different is the fact that there is no overlap in the elements that the two models find relevant and compare in order to determine similarity between the concepts.

On the other hand, we would like to stress one important similarity. Both the LSA and *contrast model* fail to capture structural relations. Although LSA looks at concepts in context of other concepts, the model is fed by simple co-occurrences of words, without information about specific relations between the concept of interest and its neighboring concepts. Analogous to how Tversky's *contrast model* assumes that concepts can be compared as mere lists of intrinsic properties, LSA assumes that concepts can be compared as lists of extrinsic properties. While we admit that the fact that subtle semantic relations could be extracted from so little information (pure co-occurrence statistics) is one of the most impressive properties of the LSA model, at the same time its simplicity also puts limits on its validity as a model of human cognition.

Just as two entities made of same parts could be very different depending on how these parts are combined, same applies for combinations of concepts - concepts that occur in similar contexts may be related to their neighboring concepts in different ways. As we already know, human similarity judgments are sensitive to the relational structure of intrinsic properties of items (Markman & Gentner, 1993), thus it is reasonable to expect they should also be sensitive to the relational structure of extrinsic properties. This has been confirmed by a recent study that demonstrated that perceived similarity is affected both by concepts playing the same role (e.g. predator) and concepts playing complementary roles within the same relational systems (e.g. predator vs. prey) (Jones & Love, 2007). Importantly, the same study also showed that same-role relation (e.g. predator - predator) affects perceived similarity more than complementary relation (e.g. predator - prey) (Jones & Love, 2007). This is an effect that could not be predicted by the LSA model as the model does not account for structural relations. To take a simplified example, suppose we were to compare the similarities between items B and C and B and D. Based on our experience, we know that these items are related in the following manner: "A eats B."; "A eats C."; "A is eaten by D". Since LSA only codes co-occurrences of items, it would fail to capture the difference in similarity between B - C (prey - prey) and B - D pairs (prey - predator).

Love and Jones (2007) suggested a model that follows the logic and underlying principles of LSA (Landauer & Dumais, 1997) but improves upon the previous model by being sensitive to sentence structure. While in LSA sentences were represented as sets of words, in ROLES (Relations Offer Latently Extracted Similarities) sentences are represented as trees that encode relational information between the words. Given that, ROLES could predict differences between same and complementary role effects (Love & Jones, 2007). In regard to their assumptions of effects of relational structures, difference between LSA (Landauer & Dumais, 1997) and ROLES (Love & Jones, 2007) is similar to the difference between the contrast model (Tversky, 1977) and structural alignment model (Markman & Gentner, 1993).

Although LSA (Landauer & Dumais, 1997) cannot predict any difference between the way that the same role and complementary role affect perceived similarity, there is an interesting and surprising pattern of results that can be accounted by LSA. It has been reported that based on the LSA model, coffee is more similar to cup than to tea; cow is more similar to barn than to pig; and car is more similar to driver than to motorcycle (Love and Jones, 2007). In other words, some word pairs denote concepts that share features and belong to the same semantic category (e.g. cow – donkey), but are estimated as less similar than concepts that do not share features but are linked based on their complementary roles (e.g. hammer – nail). Based on both the *structural alignment* model (Markman & Gentner, 1993) and the *contrast* model (Tversky, 1977), we would predict that *cow* and a *donkey* would be judged to be more similar than *cow* and *milk.* On one hand, similarity between *cow* and *donkey* is high because: (a) they share many common intrinsic properties (*contrast model*) and (b) because these common intrinsic properties are related in the same way in both objects (*structural alignment model*). On the other hand, the similarity between *cow* and *milk* is low since they do not share many common intrinsic features, though cows produce milk, and they do not share intrinsic relational structures. Although hardly anyone would dispute that cow is more similar to donkey than to milk, human similarity judgments run contrary to the prediction of the intrinsic features models and common sense: items that do not share intrinsic features but are linked by complementary extrinsic relations receive high similarity ratings (Wisniewski and Bassok, 1999; Estes, 2003).

In study of Wisniewski and Bassok (1999), high similarity rates for *man-tie* pair were explained by appeal to the fact that man and tie are similar, since a man might wear a tie. Although this answer might sound child-like, the participants in their study were university students. Wisniewski and Bassok (1999) offered an explanation for the unexpected effect of external relations on similarity judgments in their *dual process model of similarity*. They suggested that different types of concepts may be compatible with different types of processing. While concepts that share features are more compatible with the comparison process, pairs that are not similar but are complementary with respect to role are more compatible with the process of integration. They further proposed that concepts compatibility can override the task required processing. Thus, although the task was to compare two objects, participants tended to see the objects as complementary rather than contrasting them and this affected their similarity judgment through a process of integration. Their *dual process model of similarity* (Wisniewski & Bassok, 1999) assumes that whenever one judges the similarity of two objects, this judgment is going to be affected by the strength of extrinsic relatedness between the concepts of interest.

Assumption of the *dual process model of similarity* (Wisniewski & Bassok, 1999) that concept compatibility can override task demands is supported by other studies. These studies suggest that processing of relations between the objects happens automatically, possibly even during the object recognition phase (Sachs, Weis, Zellagui, Huber, Zvyagintsev, Mathiak, & Kircher, 2008; Estes et al, 2011; Murphy, 2002). Studies of semantic priming have shown both behavioral and neural effects of relational information. Robust facilitation effects were found when a concept was preceded by a concept (prime) that was linked by an extrinsic relation, like relation between bow and arrow, or hammer and nail (Moss et al, 1995; Chwilla & Kolk, 2005; Estes & Jones, 2009; Estes et al, 2011). Based on the results of priming studies it has been suggested that both information about intrinsic and extrinsic properties of items is automatically available when a concept is activated (e.g. Sachs, Weis, Zellagui, Huber, Zvyagintsev, Mathiak, & Kircher, 2008; Mirman, Walker & Graziano, 2009; Kalénine, Mirman, Middleton, & Buxbaum, 2012). Additionally, it has been demonstrated that when asked to list features of given concepts, participants tend to include significant proportion of extrinsic features (Barr & Caplan, 1987). McRae, Cree, Seidenberg and McNorgan (2005) reported that even 40% of properties listed were relational, extrinsic properties.

Since many of these findings run counter to our intuition and violate assumptions of traditional models of similarity and the hierarchical accounts of conceptual knowledge organization, there was a huge resistance to accepting the importance of extrinsic information. It was hard to understand their role in the bigger picture since they were interpreted through eyes of theories they could not possibly fit in. Higher similarity estimates for complementary pairs that do not share features (cow - barn) than for pairs of the same semantic category that share many features (cow-pig) were taken as a glitch of the LSA model (Landauer & Dumais, 1997). When similar pattern was found in human similarity judgments (Wisniewski & Bassok, 1999), it was interpreted as an artefact of the task or indicator of intrusiveness of relational information on similarity judgment. Thus, even when it was acknowledged that extrinsic relational information plays role in similarity assessment, relational information was interpreted as an intruder.

5. General discussion

In this paper, we aimed to review empirical findings and models of similarity that illustrate development of different views on one of the biggest questions in cognitive psychology: What makes for conceptual similarity? Decades of research reveals that the human sense of similarity is a highly sophisticated tool. This research also resulted in a number of models of similarity that deepened our understanding of cognitive mechanisms underlying our sense of similarity. However, these models come from different schools of thought and were based on very different views of conceptual knowledge. Consequently their assumptions regarding what determines similarity are very different.

In Figure 1, we classified models of similarity presented in this paper based on two criteria: features and form The first criteria discriminates between the models that compute similarity analyzing intrinsic features and models that base similarity estimate on analyses of extrinsic features. The second criteria distinguishes between models that assume that features can be analyzed in isolation, and those that assume that features can only be analyzed as interconnected with other elements. Note that this classification is not explicitly articulated in the cognitive psychology literature. Instead, it is intended to provide a helpful way to organize the commonalties and differences between the approaches.

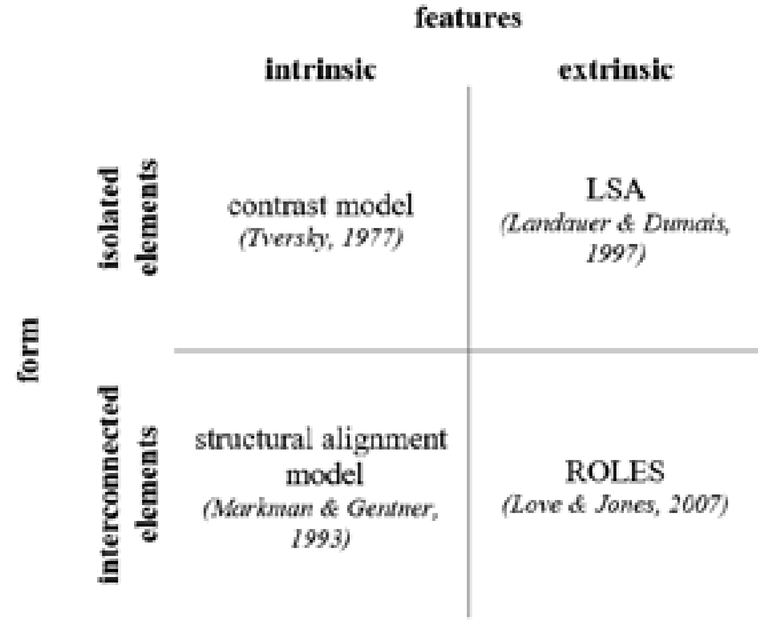

Figure 1: Two-dimensional classification of similarity models

We started our review with the model that offers formalization of the intuitive understanding of similarity as a degree of overlap in intrinsic properties of the entities – Tversky's *contrast model*. In our classification, Tversky's model takes upper left corner because it assumes that (a) entities can be analyzed as sets of independent, isolated properties and that (b) similarity is determined by entities' intrinsic properties. This view of similarity has proven to be oversimplified, as it failed to explain patterns in human similarity judgments that are sensitive to relational structure of entities' properties (e.g. Rattermann & Gentner, 1987; Markman & Gentner, 1993). A person wearing dotted shirt and striped pants is not perceived to be dressed the same way as a person wearing striped shirt and dotted pants. The *structural alignment model* (Markman & Gentner, 1993) improved upon the *contrast model* by acknowledging the role of the relational structure of intrinsic properties. This model takes bottom left cell (Figure 1) since it computes similarity based on intrinsic properties. But, in contrast to Tversky's model, that neglected role of relations between the properties, this model emphasizes the key role of the relational structure, as the overall relational information constrains what properties matter for determining similarity.

The assumption that the similarity estimate is based solely on intrinsic features analyses fits with the view of conceptual knowledge as being context free. This idea dominated psychological theories of conceptual knowledge development and organization (Quillian, 1968; Collins & Loftus, 1975; Rosch, 1975; Rogers & McClelland, 2004). If concepts encode only information about intrinsic properties, a similarity judgment between the two concepts should not be affected by information about how these concepts are related to other concepts. However, the importance of information about relations among concepts, or knowledge of concepts in the context of other concepts, has recently become prominent in cognitive psychology (Murphy, 2002; Sachs et al, 2008; Mirman, Walker & Graziano, 2009; Estes et al, 2011; Kalénine, Mirman, Middleton, & Buxbaum, 2012). This shifts the focus of the field towards analyzing the roles of intrinsic and extrinsic information in conceptual knowledge.

Given that real world objects exist in relations with other objects, it is reasonable to think their mental representations cannot be analyzed in isolation but only in context of the network of their connections with other concepts (see Murphy, 2002). A stronger version of this claim would suggest that concepts have no meaning or that they carry different meaning isolated from the network. Although LSA (Landauer & Dumais, 1997) and ROLES (Love & Jones, 2007) models were not developed to directly address this criticism, they accurately capture the gist of the idea of "concepts in context". These two models take the right side of the diagram (Fig 1). Both of these models assume that the sense of similarity does not need to be based on intrinsic features. Quite the opposite, similarity between the concepts is based on their extrinsic features. Two concepts are similar if there are related to a similar set of other concepts (LSA; Landauer & Dumais, 1997) or if they are related in the same way to other concepts (ROLES; Love & Jones, 2007). Since both LSA and ROLES are based solely on statics extracted from language (co-occurrence of words, words' distributions), their representations do not encode any information about intrinsic properties of concepts.

Thus, there is a bold line between the models on the left and the right side of the diagram. The concept dog, as viewed by models on the left, does not encode that dog chases cats. For these models, that information is not important for judgment of similarity between dog and other concepts. The concept dog, as viewed by models on the right, does not encode that dog has 4 legs, tail, etc., and this intrinsic information does not affect estimates of similarity[1]. Despite the fact that there is no overlap in elements of analyses they base estimates of similarity on, they are both able to predict patterns of human similarity judgments or patterns of behavior supported by mechanisms of similarity estimates (Markman & Gentner, 1993; Landauer & Dumais, 1997). Hence, even though these models have diametrically opposite views of conceptual knowledge and approaches to determining conceptual similarity, they might both be correct.

One possible interpretation could be that these models are measuring different aspects or different types of similarity. As it was earlier suggested by some authors (e.g. Murphy, 2002) there may not be one similarity, but rather there are different types of relations between the concepts that we label by the same term – similarity. This is an interesting proposal, which may imply that there are separate cognitive mechanisms that affect the human sense of similarity. The *dual process model* of similarity (Wisniewski and Bassok, 1999) that we described earlier illustrates one of the ideas in this direction. On the dual process model, one could assume that different aspects of conceptual similarity are useful in different cognitive tasks. Different aspects of conceptual similarity may even be useful during the same task but processing different kinds of information, which have their unique contribu-

1 Note that authors of these models (Landauer & Dumais, 1997; Jones & Love, 2007) do not claim that human conceps do not capture intrinsic information, but models themselves are restricted to operate using solely extrinsic information.

tions to the overall conceptual processing (Murphy, 2002; Simmons & Estes, 2006; Jones & Golonka, 2012). Some of the recent neurocognitive findings support this view suggesting that processing of intrinsic and extrinsic relations, or in terms of these studies taxonomic and thematic relations, engage distinct neural processes, but that both types of information are spontaneously activated during processing conceptual similarity (Chen et al, 2013).

If we would assume that intrinsic and extrinsic information analyses are addressing different kinds of similarity, it would be hard to explain findings of the studies that reported that human judgments of concepts' feature similarity and LSA estimates of conceptual similarity are significantly correlated (Simmons & Estes, 2006; Jones & Golonka, 2012). While the size of the correlation effects is moderate, this finding is very interesting having in mind that these measures are based on different sources of information, and possible distinct neural mechanisms. Yet, these correlations are surprising only if we think of intrinsic and extrinsic features of objects as independent. This is typically not the case. Objects that share intrinsic features typically also share extrinsic features. Similarity between intrinsic features makes objects compatible for interactions with similar sets of other objects. The dependency of the inside and outside structures of features makes the connection between intrinsic and extrinsic features often overlap. Taking this perspective we could expect that both models on the left and on the right side of the diagram (Figure 1) could govern our sense of similarity towards the same (or at least comparable) similarity judgment. Although correlations between different similarity estimates do not deny that these processes may require different neural resources, they do suggest that they are not necessarily addressing different faces or types of similarity. More support comes from a recent computational work that shows that hierarchical organization and category structure may arise from information about word distribution (Huebner & Willits, 2017). We would like to suggest that despite all the differences between intrinsic and extrinsic information processing, both empirical findings and computational work imply that the two keep close company. We believe that the future of studies of conceptual similarity is to determine how intrinsic and extrinsic information processing complement each other.

Hardly anyone would argue that human concepts do not code information about intrinsic features of objects, although it may not be a default for all kinds of concepts. However, it is possible that through learning, either early in development or as an adult novice, we may not have the capacity to process intrinsic information. In those cases, we could rely on extrinsic information to govern our sense of similarity. Thus we can establish a link between donkey and horse based on the overlap of the contexts they appear in, which parallels the way we would establish that link between them based on overlap in features. Although computational models of similarity are typically based on purely linguistic input, there is nothing in their basic assumptions that would suggest that the same basic principles could not govern our processing of non-linguistic information. Thus we could know that two objects are similar even before we ever perceptually experience them or before we are able to analyze their intrinsic properties. In this example, sense of similarity based on extrinsic features overlap could aid or guide our sense of similarity based on intrinsic features. Following the gist of the idea of the structural alignment model where prior to feature comparison process, features first need to be aligned, put in correspondence, in order to constrain the comparison process; extrinsic features overlap may constrain intrinsic features comparison. That way, two senses of similarity or two levels of analyses may be complementary as dual aspects of the same psychological process.

This critical review aimed to draw parallels between two streams in literature on conceptual similarity, each focusing on the contribution of one of two types of information: intrinsic and extrinsic features. We believe that taken together, the empirical findings we presented here support the hypothesis that both intrinsic and extrinsic information determines human similarity judgments. Although effects of processing of overlap in intrinsic and overall in extrinsic features were typically studied independently, the aim of this review was to show not only that combining insights from different approaches to similarity may be a fruitful approach, but also that without such an approach, our understanding of what determines human sense of similarity would stay incomplete.

Olivera Savić
Department of Psychology, The Ohio State University, United States

Hope Sample
Department of Philosophy, The Ohio State University, United States

Literature:

Barr, R. A., & Caplan, L. J. (1987). Category representations and their implications for category structure. *Memory & Cognition*, 15, 397–418.

Bassok, M., & Medin, D. L. (1997). Birds of a feather flock together: Similarity judgments with semantically rich stimuli. *Journal of Memory and Language*, 36, 311-336.

Chen, Q., Lei, Y., Li, P., Xi, L., Li, F., & Li, H. (2013). How do taxonomic versus thematic relations impact similarity and difference judgments? An ERP study. *International Journal of Psychophysiology*, 90(2), 135–142.

Chwilla, D. J. & Kolk, H. H. J. (2005). Accessing world knowledge: Evidence from N400 and reaction time priming. *Cognitive Brain Research*, 25, 589 – 606.

Collins, A.M., & Loftus, E.F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407-428.

Estes,W. (1994). *Classification and cognition*. Oxford University Press.

Estes, Z., Golonka, S., & Jones, L. L. (2011). Thematic thinking: The apprehension and consequences of thematic relations. *Psychology of Learning and Motivation*, 54:249–294.

Estes , Z. & Jones , L. L. (2009). Integrative priming occurs rapidly and uncontrollably during lexical processing. *Journal of Experimental Psychology: General*, *138*, 112 – 30.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2), 155–170.

Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science*, 6, 298–304.

Jones, M., & Love, B. C. (2007). Beyond common features: The role of roles in determining similarity. *Cognitive Psychology*, 55, 196–231.

Kalénine S., Mirman D., Middleton E.L., & Buxbaum L.J. (2012). Temporal dynamics of activation of thematic and functional knowledge during conceptual processing of manipulable artifacts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38 (5), 1274-1295.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104, 211-240.

Lin, E. L. & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General*, 130, 3-28.

Markman, A. B., & Gentner, D. (1993). All differences are not created equal: A structural alignment view of similarity. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, 682-686.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavioral Research Methods, Instruments, and Computers*, 37, 547–559.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1990). Similarity involving attributes and relations: Judgments of similarity and difference are not inverses. *Psychological Science*, 1, 64-69.

Medin, D. L. & Rips, L. J. (2005). Concepts and categories: memory, meaning, and metaphysics. In K. Holyoak & R. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 37–72). Cambridge University Press, Cambridge, UK.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207- 238.

Mirman, D., Walker, G.M., & Graziano, K.M. (2011) A Tale of Two Semantic Systems: Taxonomic and Thematic Knowledge, 2211-2216. *In Proceedings of the 33rd Annual Cognitive Science Society Meeting.*

Moss, H.E., Ostrin, R.K., Tyler, L.K., & Marslen-Wilson, W.D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 863-883.

Murphy, G. L. (2001). Causes of taxonomic sorting by adults: A test of the hematic-to- taxonomic shift. *Psychonomic Bulletin & Review*, 8, 834-839.

Murphy, G. L. (2002). *The Big Book of Concepts*, MIT Press.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 104-114.

Quillian, M. R. (1968). *Semantic memory*. In M. Minsky (Ed.), Semantic information processing. Cambridge, MA: MIT Press.

Quine, W. V. (1969). *Natural kinds*. In W. V. Quine (Ed.), Ontological relativity and other essays (pp. 114-138). New York: Columbia University Press.
Rattermann, M. J., & Gentner, D. (1987). Analogy and similarity: Determinants of accessibility and inferential soundness. In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society* (pp. 23-35). Lawrence Erlbaum Associates, Mahwah, JN.
Rogers, T. T., &McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA:MIT Press.
Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192-233.
Sachs, O., Weis, S., Krings, T., Huber, W., & Kircher, T. (2008). Categorical and thematic knowledge representation in the brain: Neural correlates of taxonomic and thematic conceptual relations. *Neuropsychologia*, 46, 409-418.
Sachs, O., Weis, S., Zellagui, N., Huber, W., Zvyagintsev, M., Mathiak, K., & Kircher, T. (2008). Automatic processing of semantic relations in fMRI: neural activation during semantic priming of taxonomic and thematic categories. *Brain Research*, *1218*(2), 194–205.
Saussure, F. (1959). Course in general linguistics (W. Baskin, Trans.). New York: McGraw-Hill. (Original work published 1915)
Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review*, 96, 125–144.
Tulving (1972). *Episodic and semantic memory*. In E. Tulving & W. Donaldson (Eds.), Organization of Memory (pp. 381-403). New York: Academic Press.
Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Erlbaum.
Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352.
James , W. (1890). *The Principles of Psychology* . Cambridge, MA : Harvard University Press .
Wisniewski, E. J. & Bassok, M. (1999). What makes a man similar to a tie? Stimulus compatibility with comparison and integration. *Cognitive Psychology*, 39, 208-238.

Šta predstavlja pojmovna sličnost?
(Apstrakt)

U psihologiji, značajan broj fenomena objašnjava se sličnošću. Dok se ovaj termin pokazao kao koristan pri razumevanju i definisanju drugih enomena, sličnost sama po sebi jako je slabo shvaćena i definisana. Ovde nameravamo da diskutujemo o trenutnom statusu pojma sličnosti, onako kako se on upotrebljava u istraživanjima u oblasti kognitivne psihologije i kognitivne neuronauke.

Ključne reči: Pojmovna sličnost, model strukturalnog poravnanja, latentna semantička analiza