
Melodic Cue Abstraction, Similarity, and Category Formation: A Formal Model

EMILIOS CAMBOUROPOULOS

Austrian Research Institute for Artificial Intelligence, Vienna

In the first part of this article, the notions of identity, similarity, categorization, and feature salience are explored; musical examples are provided at various stages of the discussion. Then, formal working definitions are proposed that inextricably bind these concepts together. These definitions readily lend themselves to the development of a formal model for clustering—the *Unscramble* algorithm—which, given a set of objects and an initial set of properties, generates a range of plausible categorizations for a given context. Finally, as a test case, the clustering algorithm is used to organize a number of melodic segments, taken from a monophonic piece by J. S. Bach, into motivic categories; the algorithm also determines a prototype for each cluster and uses these prototypical descriptions for membership prediction tasks. The results of the computational system are compared with the empirical results obtained for the same data in two earlier studies (I. Deliège, 1996, 1997).

Στο πρώτο μέρος αυτής της μελέτης γίνεται συνοπτική διερεύνηση, με αναφορά σε μουσικά παραδείγματα, των εννοιών της ταυτοσημίας, της ομοιότητας και της κατηγοριοποίησης. Στην συνέχεια, δίνονται τεχνικοί ορισμοί των εννοιών αυτών οι οποίοι αναδεικνύουν την δυναμική αλληλεξάρτησή τους. Αυτοί οι ορισμοί αποτελούν τη βάση στην οποία στηρίζεται η ανάπτυξη του αλγόριθμου κατηγοριοποίησης *Unscramble*, ο οποίος οργανώνει σε κατηγορίες ένα σύνολο αντικειμένων ενώ παράλληλα αναδεικνύει τις πιο χαρακτηριστικές ιδιότητες που περιγράφουν την κάθε κατηγορία. Τέλος, ο προτεινόμενος αλγόριθμος χρησιμοποιείται για την οργάνωση μελωδικών αποσπασμάτων, προερχόμενων από ένα μονοφωνικό κομμάτι του Ι.Σ.Μπαχ, σε μοτιβικές κατηγορίες, και γίνεται σύγκριση των αποτελεσμάτων της αλγοριθμικής ταξινόμησης με τα αποτελέσματα δύο προγενέστερων συναφών εμπειρικών μελετών (Deliège, 1996, 1997).

ONE significant component of musical understanding is the ability of listeners to cluster musical materials together into categories such as motives, themes, and so on. Salient musical cues enable listeners to make similarity judgments between various musical materials and to organize these into meaningful groups. In this study, the notions of feature salience,

Address correspondence to Emilios Cambouropoulos, Austrian Research Institute for Artificial Intelligence, Schottengasse 3, A-1010, Vienna, Austria. (e-mail: emilios@ai.univie.ac.at)

ISSN: 0730-7829. Send requests for permission to reprint to Rights and Permissions, University of California Press, 2000 Center St., Ste. 303, Berkeley, CA 94704-1223.

similarity, and categorization are discussed, and then working formal definitions are presented. It is herein maintained that cue abstraction, similarity judgments, and categorization processes are inextricably bound together in such a way that each of these can be defined only in relation to the others.

A formal model for clustering has been developed on the basis of these definitions. Given a segmentation of a melodic surface and an initial representation of each segment in terms of a number of attributes (these reflect melodic and rhythmic aspects of the segment at the surface and at various abstract levels), the *Unscramble* algorithm organizes these segments into “meaningful” categories. The proposed clustering algorithm automatically determines an appropriate number of clusters and also the characteristic (or defining) attributes of each category. A limited number of attempts have been made to use clustering techniques to organize melodic segments into motivic categories. A brief survey and comparison of some existing algorithms is presented in Hötheker, Hörnel, and Anagnostopoulou (2001).

A number of psychological studies have attempted to examine the notions of melodic similarity, categorization, and cue abstraction by using real melodic material (e.g., Carterette, Hohl, & Pitt, 1986; Deliège, 2001; Eerola, Järvinen, Louhivuori, & Toiviainen, 2001; Koniari, Predazzer, & Mélen, 2001; Lamont & Dibben, 1997, 2001; Mélen & Wachsmann, 2001; Pollard-Gott, 1983). The most extended studies, however, have been performed by Deliège (see overviews in Deliège, 1997, 2001, and Deliège & Mélen, 1997), wherein issues of feature salience (*cue abstraction*), musical similarity, and prototypical description of categories (*imprint formation*) in musical listening are empirically examined.

It is interesting to compare the performance of a computational model against the results given in empirical studies. A computational approach requires explicit representations of the musical materials and detailed formal descriptions of similarity and categorization processes. The various processes can thus be traced and analyzed step-by-step in a way that usually is not possible in empirical studies.

This study attempts to replicate, by means of computational modeling, two psychological experiments on cue abstraction and categorization performed on a monophonic piece by J. S. Bach (Deliège, 1996, 1997). The results of the computational approach are compared with the empirical results, and convergences and deviations are reported. The clusters produced by the algorithm correspond closely to the categories provided in the empirical study. The application of the algorithm confirms most of the suggestions presented in the psychological studies regarding which cues play the most significant role in categorization tasks and how similar melodic segments are organized into categories.

In the following sections, initially the concepts of identity, similarity, and categorization are discussed. Then, the *Unscramble* algorithm is described. Finally, results of the application of the algorithm on motivic segments of J. S. Bach's *Allegro Assai* of the Sonata for Solo Violin in C major BWV 1005 is presented, and various interesting aspects of the computational experiment are discussed.

Theoretical Concepts

IDENTITY

Before attempting to describe the notion of similarity, it is important to discuss briefly the principle of identity and to try to clarify its usage within this text. Without getting into a deep ontological discussion, an *entity* is herein taken to refer to a complete and distinct thing—concrete or abstract—such as an object, an event, a structure, a function, a goal, and so on (e.g., a pencil, a robin, a song, an emotion, an action such as running or sleeping). A *property* is any predicate that may be used to describe an entity.

Is it possible for two different entities (that have different spatiotemporal properties) to be identical? For example, is it possible that two drops of water or two middle-C notes played on the same instrument may be identical? Leibniz's response to such a question would be that "there are no two individuals indiscernible from one another" (Fourth paper to Clarke, Sec. 4, quoted in Stroll, 1967, p. 122) or "there are not in nature two individuals indiscernible from one another" (G. VII. 393 (D. 258) in *Extracts from Leibniz* in Russell, 1949, p. 219). This principle is referred to as the *principle of the identity of indiscernibles*. Stroll (1967) states that "Leibniz's language suggests that he considered this principle to be an empirical law; that if we were to find two items (say, two drops of water) apparently possessing exactly the same set of internal features, further investigation (by means of a microscope, for instance) would show that they differed from one another." (p. 122). He then continues: "But reflection upon [Leibniz's] use of the expressions 'intrinsic quality' and 'internal difference' suggests that he covertly employed the principle as if it were a logical truth, to which no empirical finding would be a counter-instance." (p. 122). Many philosophers have rejected this principle when presented as logically necessary (Black, 1952), but it is accepted when seen as an empirical law.

This principle is connected, according to Russell, to Leibniz's implied assertion "that every substance has an infinite number of predicates"

(Russell, 1949, p. 60). “[I]ndividuality involves infinity, and only he who is capable of understanding it [infinity] can have knowledge of the principle of individuation of such or such a thing” (G.V. 268 [N.E. 309] in Extracts from Leibniz in Russell, 1949).

According to Russell’s definition of identity, two entities x and y are identical if and only if the same properties (predicates) are satisfied by both (Russell & Whitehead, 1964, vol. i, def. 13.01). The identity relation is an equivalence relation, that is, it is reflexive, symmetric, and transitive. But, is this definition of any use if two entities have an infinite number of properties? How is it that one says that two different drops of water or two middle-C notes are identical?

The key to answering these questions is that two entities are judged identical only when a finite number of properties *that are considered salient* for a given domain of discourse are demarcated. When we say that two objects are identical we mean that all the properties (predicates) that describe the two objects—taken from a set of predefined properties that are considered to be pertinent in a given context—have the same values. Quine (1950) emphasizes the importance of a domain of discourse: “In general we might propound this maxim of the *identification of indiscernibles*: Objects indistinguishable from one another within the terms of a given discourse should be construed as identical for that discourse.” He continues that this maxim “is relative to a discourse, and hence vague in so far as the cleavage between discourses is vague. It applies best when the discourse is neatly closed, like the propositional calculus; but discourse generally departmentalizes itself to some degree, and this degree will tend to determine where and to what degree it may prove convenient to invoke the maxim of identification of indiscernibles” (p. 626).

The most crucial factor in establishing “meaningful” identities is selecting the set of properties that are pertinent in describing a set of entities in a given situation. This set of properties is not absolute but depends on the task at hand. For instance, two tunes may be most commonly considered identical in the Western tradition if they both are composed of the same sequence of diatonic pitch intervals and duration ratios, that is, the same musical surface. If, on the other hand, in a different domain their expressive or spectrographic properties are considered to be most pertinent, then they may be judged as being nonidentical.

SIMILARITY

Similarity is a difficult and obscure notion. How does it relate to identity? What are the conditions and limits under which two entities may be considered similar? For a given set of pertinent properties and following from Russell’s definition of identity, similarity is very often defined as *par-*

tial identity, that is, two entities are similar if they have *some* properties (predicates) that are the same but not necessarily all. Pairs of entities may be compared, and one pair may be judged as being more similar than another if its members share more common properties than the members of the other pair.

Similarity between two entities may be calculated by simply counting the number of matches between their properties. Alternatively, similarity may be defined as a function of the differences between all the pairs of properties these objects possess. For example, according to the traditional multidimensional scaling model (Shepard, 1962a, 1962b), similarity between objects x and y is a monotonic decreasing function f of interpoint distance:

$$s(x,y) = f[d(x,y)]$$

where $s(x,y)$ is a similarity rating between x and y , $d(x,y)$ is the distance between the two points of the objects' attribute vectors in a multidimensional attribute space. (For a brief summary of commonly used metrics, see Murtagh, 1993, pp. 228–230).

If all properties receive equal weights for the metric $d(x,y)$, then this definition of similarity is equivalent (for binary features and Hamming distances) to the former definition (i.e., partial matching of properties). If, on the other hand, properties are given different weights reflecting the intuition that not all properties are equally important for a given object, then there is a significant departure from the former traditional definition of similarity. For instance, the members of a pair of objects that have in common only one important property may be judged as being more similar than the members of another pair that share two or more less salient properties.

The similarity definitions just given imply that the similarity relation is reflexive and symmetric but not transitive. There exist, however, other models that allow asymmetric definitions of similarity. For example, Tversky (1977) proposed that similarity between two entities may be defined as a function of their common properties minus the properties that are distinctive to either of them:

$$s(x,y) = \theta \cdot f(X \cap Y) - \alpha \cdot f(X - Y) - \beta \cdot f(Y - X),$$

where $s(x,y)$ is the similarity between two objects; X and Y are the feature sets of x and y , respectively; and θ , α , and β are parameters that are used to reflect prominence of common and distinctive features. Tversky's model of similarity has been proved to be very useful in describing (empirically) observed similarities but is rather impractical if used to predict similarities between entities because it requires a very elaborate representation of each individual entity. That is, the model requires that the individual sets of all

the features that are important for the description of each object be precisely defined (rather than using only one general set of features that accounts for all the objects) and/or all three parameters θ , α , and β be given in advance for each ordered pair of objects. Tversky's model fails to address the question of how people determine which properties are relevant for a similarity comparison (see Barsalou, 1992, pp. 282–284).

Alternatively, Krumhansl (1978) proposes an extension of the multidimensional similarity definition, namely the distance-density model, that accounts for asymmetric judgments and contextual aspects of similarity. The distance-density model is based on the assumption that “two points in a relatively dense region of a stimulus space would have a smaller similarity measure than two points of equal interpoint distance but located in a less dense region of the space” (Krumhansl, 1978, p. 446). According to this model, the distance $d(x,y)$ in the similarity function of the multidimensional scaling model— $s(x,y) = f[d(x,y)]$ —is replaced by a modified distance function $d'(x,y)$: $d'(x,y) = d(x,y) + \alpha \cdot \delta(x) + \beta \cdot \delta(y)$, where $d(x,y)$ is the interpoint distance, $\delta(x)$ and $\delta(y)$ are measures of spatial density in the neighborhoods of x and y , and α and β are constants that reflect the relative weight given to the densities $\delta(x)$ and $\delta(y)$. For instance, “if $\alpha < \beta$, then $s(x,y) > s(y,x)$ if and only if $\delta(x) < \delta(y)$, that is, in directional similarity tasks, asymmetries would be expected to be associated with differences in the densities in the regions surrounding the two points in the geometric configuration” (Krumhansl, 1978, p. 453). This definition of similarity augments the traditional definition by incorporating a density factor that relies on local context.

A common characteristic of all the above definitions is that none of them incorporates a notion of categorization. These definitions of similarity (usually the symmetric ones) are commonly used as prerequisites for other categorization models that predict possible clusterings of objects but they are not explicitly linked to a notion of categorization.

CATEGORIZATION

In the course of this text, the word *category* will be taken to refer to a set of entities that are grouped together on the basis of some criteria. The conditions for classification are commonly referred to as the *intention* of a concept and the set of entities that are members of a category the *extension* of the concept. The term *concept* “refers to the idea or notion by which an intelligence is able to understand some aspect of the world” (Hampton & Dubois, 1993, p. 13).

According to the classical *monothetic* definition, a category is constituted of all the entities that possess a set of properties or satisfy a set of conditions (see Sutcliffe, 1993). Most commonly, these conditions are taken to be singly necessary and jointly sufficient.

A different approach to formalizing the notion of categories has emerged following Wittgenstein's approach to the notion of "family" and "family resemblance" (Wittgenstein, 1953). According to the polythetic view, a category consists of individuals that have a large number of properties from a given set P , and each property is possessed by a large number of members but no property is possessed by all the members of the category (Beckner, 1959, p. 21). The problem with this definition is to determine when a "large number" is large enough, that is, to define a limit above which entities share enough properties so as to be considered members of a category.

The polythetic definition of categories underlies prototype models (Rosch, 1975; see Hampton, 1993 for an overview) and exemplar models of categorization (Estes, 1994). According to the prototype view, members of a category are determined by their similarity to the category's prototype and "a prototype concept is constituted by a *set of attributes with associated values* (where a particular attribute-value pair corresponds to a property), each with a particular weight corresponding to its 'definingness' or contribution to the concept's definition." (Hampton, 1993, p. 73). Membership and typicality of an instance are judged in relation to a similarity measurement of the individual to the category's prototype (i.e., the weighted attribute-value set), or to the exemplars for exemplar models. There exists a criterion on the similarity scale over which individuals are considered to be members of the category and their typicality is proportional to the similarity rating (i.e., the higher the rating for an instance, the higher its typicality).

Prototype models account for many phenomena observed in the way humans make categorizations in everyday situations, such as flexibility of category boundaries, gradedness and typicality of members, ambiguity of membership (e.g., is a tomato a fruit or a vegetable?), and so on.

The prototype of a concept and the similarity criterion can be determined by direct experimentation and then used for further predictions. If one, however, wants to determine prototypes and similarity criteria for a set of entities so that these entities may be organized into categories, then the prototype definition of a category reveals its weaknesses. How can one discover a relevant similarity threshold for determining the membership of objects in a category if the prototype is not known? How can the prototype (i.e., a weighted set of characteristic attribute values) be determined? If the extension of a category is given, then a prototype can be defined (by finding the most characteristic properties that are possessed by most members), but that means that one knows in advance the category members. But how could the category be determined without reference to the prototype when it is defined in terms of the prototype? Sutcliffe remarks that "there must first be a family before one can observe any family resemblances, and thus one cannot define a family by reference to family resemblances!" (Sutcliffe, 1993, p. 46).

The prototype view on categorization relies to some extent on either some form of independent bottom-up, data-driven, clustering-like analysis (see Mechelen et al., 1993, part II) or on top-down theory-based approaches (e.g., Murphy, 1993; Murphy & Medin, 1985) or a mixture of the two.

Both of the preceding descriptions of categories can accommodate conjunctive as well as disjunctive intensional descriptions (especially for monothetic categories, Sutcliffe [1993, p. 59] argues that disjunctive concepts have a sound logical basis). It is asserted herein that disjunctive concepts are hard to work with when dealing with unsupervised category formation tasks. The reason for this assertion is that the space of all possible conjunctive descriptions (for a given set of entities) through which a search has to be pursued is augmented explosively if disjunctive concepts are considered as well. If instances, though, of a category are known in advance — as in supervised learning — then disjunctive descriptions may be convenient (e.g., if “couples” are represented on an instance space by the “color” of each partner, then categories such as “mixed couples” and “same-color couples” are not possible unless either disjunctive concepts are accommodated or the initial representation is altered).

The debate between the “classical” and the “modern” view is heated. Hampton (1993, p.76) argues that “classical monothetic concepts can be treated as special cases of prototype models in which the membership criterion has been set very high on the similarity scale, so that the criterial level of similarity cannot be achieved without the core properties.” Contrastingly, Sutcliffe argues that “the ‘modern view’ developed by Rosch on the basis of Wittgenstein’s and Beckner’s notion of polythetic class, is incoherent and unworkable” (Sutcliffe, 1993, p. 62).

In this study, it is suggested that the distinction between the monothetic and polythetic views on categorization is not as sharp as many would argue (e.g., Lakoff, 1987). For instance, if an exact threshold is set for a polythetic category, then a sharp boundary is defined (some form of boundary is necessary in any case: it doesn’t make much sense to say, for instance, that a chair is a very atypical member of the category “bird” — it simply isn’t a bird). If overlapping of categories is allowed, then ambiguity and gradedness is introduced (for both monothetic categories and polythetic categories with sharp boundaries), that is, the more categories an entity belongs to, the more ambiguous it is and the less typical a member of a category it is. If the two definitions of category are dissociated from metaphysical claims and are seen simply as formal descriptions of the notion of category, then there can be only pragmatic criteria as to their usefulness and efficiency.

It is clear from the preceding discussion that all the members of a category are necessarily pairwise similar because they necessarily share some common properties (they share at least the property of belonging in the

same category!), but the converse is not necessarily true, that is, similar entities are not necessarily members of the same category. The notions of similarity and category can be brought into a close relation if a threshold is introduced in the definition of similarity (see section titled “A Working Formal Definition of Similarity and Categorization”).

Asymmetric definitions of similarity are avoided in clustering models because such definitions would make it possible for an object to be both a member and a nonmember of a category according to the two different asymmetric similarity measures (a commonly hypothesized property of a category is that it either includes or excludes a certain object).

SIMILARITY AND CATEGORIZATION BOUND TOGETHER

A commonly encountered hypothesis on which many categorization models are grounded is that categorization is strongly associated with the notion of similarity, that is, similar entities tend to be grouped together into categories.

However, there are different views on the relation between similarity and categorization (Goldstone, Medin, & Gentner, 1994; Medin, Goldstone, & Gentner, 1993). On the one hand, similarity is considered to be too flexible and unwieldy to form a basis for categorization, that is, any two entities may be viewed as being similar in some respect (e.g., a car and a dog are similar in that both weigh less than 10 tons, but these objects are not normally considered to be members of the same category!). On the other hand, similarity is regarded to be too narrow and restricting to account for the variety of human categories (e.g., a whale is considered a mammal even though it appears to be more similar to fish). Goodman (1972) doesn't hesitate to reject altogether the notion of similarity, claiming that “similarity tends under analysis either to vanish entirely or to require for its explanation just what it purports to explain” (p. 446). Rips (1989) claims that “there are factors that affect categorisation but not similarity and other factors that affect similarity but not categorisation. . . . there is a ‘double dissociation’ between categorisation and similarity, proving that one cannot be reduced to the other” (p. 23).

This debate is directly linked to a further issue; that is, how entities and their properties are represented. If objects are described in terms of mainly perceptual (e.g., visual or auditory) properties, then, obviously similarity is insufficient for many categorization tasks, whereas, if any sort of properties—perceptual or abstract or relational—are considered, then similarity becomes too flexible.

It seems that the notions of categorization, similarity, and the representation of entities/properties are strongly interrelated. It is not simply the case that one starts with an accurate description of entities and properties,

then finds pairwise similarities between them and, finally, groups the most similar ones together into categories (Figure 1a). It seems more plausible that as humans organize their knowledge of the world, they alter their representations of entities concurrently with emerging categorizations and similarity judgments (Figure 1b).

One of the main assumptions made in this study is that similarity *always* depends on context (i.e., it is contextually defined), and when similarity seems to be relatively stable, this is so simply because the context (e.g., the structure of the natural world or a specific cultural system) tends to be quite stable. Of course, there are some general perceptual constraints as to what is perceptible in the first place, but from there on different properties of entities become more prominent in a given context for a specific categorization task or for a similarity judgment. Tversky (1977) has highlighted the importance of context in similarity judgments and has shown how properties of objects become *diagnostic* within a specific context; he treats, however, these contextual effects on similarity as specific cases/exceptions rather than the norm (his definition of similarity is independent of categorization). Context dependency of musical similarity and categorization is supported by the empirical work of Lamont and Dibben (2001).

As a first general example, consider Figure 2. Which of objects *b*, *c*, and *d* is most similar to object *a*? One might, cautiously, select one of these objects or refuse to answer the question altogether. If, though, these objects are placed in a context such as a barber shop or an office or a surgical operating room, then it becomes apparent which objects are more similar and are actually categorized together, and which properties of the objects are more prominent and diagnostic in that specific context—for instance, within the context of a barber’s shop, objects *a* and *c* are more similar and they tend to be categorized together because they share barber-related properties (e.g., “hair-cutting”).

A second example from the musical domain that highlights the contextual nature of similarity and categorization is presented next. A musical work may be considered as a local context within which things like motives, themes, harmonic progression groups, and so on emerge. Trying to discover the similarity of two isolated musical passages will usually pro-

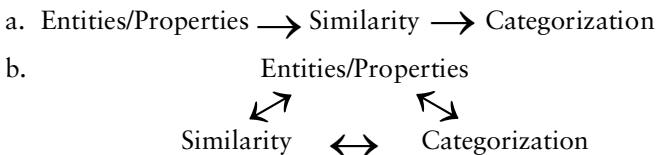


Fig. 1. Relations among entities/properties, similarity, and categorization.

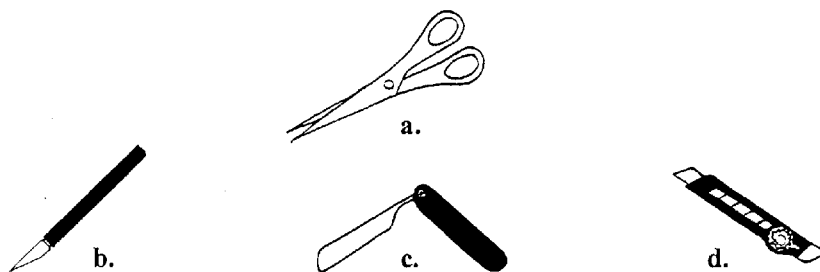


Fig. 2. Which of objects *b*, *c*, and *d* is most similar to object *a*?

duce dubious or relatively uninteresting results. Consider, for instance, the musical passages in Figure 3. In which of the two pairs are the two passages more similar? Some might select the first pair, others the second pair, and still another group might refuse to make a judgment. It is suggested that perhaps this similarity experiment is simply poorly designed in the first place, and perhaps subjects of the third group are right in refusing to make a judgment. The problem seems to be that these excerpts are taken out of their context. As it happens, the first two passages are very dissimilar—actually contrasting—within the homogeneous minimal context of Steve Reich’s *Electric Counterpoint*, whereas the second two are very similar within the very diverse context of Iannis Xenakis’ *Keren*. Context seems to be paramount in our establishing similarities and categories between musical passages, and it is asserted that it is not possible to find an absolute criterion for defining what things are similar *in general*.

As a further musical example, consider the following two rhythmic figures: ♪♪♪♪♪ and ♪♪. Are these two similar or dissimilar? Again the problem in answering this question seems to be that these excerpts are out of context. For instance, these two figures are rather dissimilar—actually contrasting—within the homogeneous rhythmic context of J. S. Bach’s *Allegro Assai* of the Sonata for Solo Violin in C major BWV 1005 (the eighth-note figure may be seen as a distinctive feature characterizing one of the two main motives; see Figure 4), whereas they are more similar within a more



Fig. 3. In which of the two pairs are the two passages more similar?

diverse melodic context in which, for instance, sixteenth-note triplets are quite common. Again context seems to play a significant role in the way we perceive similarities between musical entities.

The psychological theory of musical form proposed by Deliège (see Deliège, 1996, 2001; Deliège & Mélen, 1997) empirically examines issues of property prominence (*cue abstraction*), musical similarity, and prototypical description of categories (*imprint formation*) in musical listening. Deliège's work seems to be in line with the description of entities/properties, similarity, and categorization in the current proposal; however, the model presented in this article establishes direct formal links between these notions in a way not encountered explicitly in other cognitive accounts of musical understanding.

In the light of the preceding discussion, formal definitions of similarity and category will be given wherein the two notions are interdependent, that is, changes in similarity result in category changes, and vice versa.

The *Unscramble* Clustering Algorithm

In this section, first, formal definitions are proposed for the notions of property salience, similarity, and category, and, then, a brief description is given of the *Unscramble* clustering algorithm that was developed primarily for dealing with clustering problems in the musical domain. (A detailed description that includes also issues of computational complexity and more musical examples is given in Cambouropoulos, Smail, & Widmer, 1999, and Cambouropoulos & Widmer, 2000).

A WORKING FORMAL DEFINITION OF SIMILARITY AND CATEGORIZATION

Let T be a finite set of entities and P the union of all the sets of properties that are pertinent for the description of each entity. If $d(x,y)$ is the distance between two entities x and y , and h is a distance threshold, we define similarity $s_b(x,y)$ as follows:

$$s_b(x,y) \begin{cases} 1 \text{ iff } d(x,y) \leq h \text{ (similarity)} \\ 0 \text{ iff } d(x,y) > h \text{ (dissimilarity)} \end{cases} \quad (1)$$

In other words, two entities are *similar* if the distance between them is smaller than a given threshold and *dissimilar* if the distance is larger than this threshold.

The preceding definition of similarity is brought into a close relation with a notion of category. That is, within a given set of entities T , for a

set of properties P and a distance threshold h , a category C_k is a maximal set:

$$C_k = \{x_1, x_2, \dots, x_n | x_i \in T\} \text{ with the property:} \tag{2}$$

$$\forall i, j \in \{1, 2, \dots, n\}, s_b(x_i, x_j) = 1$$

In other words, a category C_k consists of a maximal set of entities that are pairwise similar to each other for a given threshold h .

A category, thus, is inextricably bound to the notion of similarity; all the members of a category are necessarily similar and a maximal set of similar entities defines a category. According to Definition 1, similarity is not merely the inverse of distance, but additionally requires a threshold that can be determined in relation to a specific categorization description for a given context.

As the similarity function s_b is not transitive, that is, $s_b(x, y)$ and $s_b(y, z)$ does not imply $s_b(x, z)$, the resulting categories need not be disjoint (i.e., equivalence classes). In other words, overlap between categories is permitted.

The distance threshold may take values in the range of $0 \leq h \leq d_{max}$, where the distance d_{max} is defined as the maximum distance observed between all the pairs of entities in T . For $h = 0$, every object in T is a monadic category; for $h = d_{max}$ all the objects in T define a single category.

PROPERTY SALIENCE

The salience of a certain property will be represented herein by a weight value; higher values are given to more salient properties. For a given clustering description of a set of entities, the initial weights of properties can be altered in relation to their “diagnosticity,” that is, properties that are unique to members of one category are given higher weights, whereas properties that are shared by members of one category and its complement are attenuated. A function that calculates the weight of a single property p could be:

$$w = \lfloor m/n - m'/(N - n) \rfloor, \tag{3}$$

where m = number of objects in category C_k that possess property p , m' = number of objects *not* in category C_k that possess property p (i.e., objects in $T - C_k$), n = number of objects in C_k , and N = number of objects in T . The maximum weights of each property calculated for each category are then selected for a given clustering.

In order to calculate weights according to this function (Eq. 3), category information is necessary. This way, property salience, similarity, and categorization are interrelated and linked together. The preceding definitions form the basis for the development of the *Unscramble* algorithm described in the next section.

THE UNSCRAMBLE ALGORITHM

The *Unscramble* algorithm is a clustering algorithm that, given a set of objects and an initial set of properties, generates a range of plausible clustering descriptions for a given context. During this dynamically evolving process, the initial set of properties is adjusted so that a satisfactory description is generated. There is no need to determine in advance an initial number of clusters, nor is there a need to reach a strictly well-formed (e.g., nonoverlapping) categorization. At each cycle of the process, weights are calculated for each property according to how characteristic each property is for the emergent clusters.

The input to the *Unscramble* algorithm is a set of objects each described by an m -dimensional property vector. Each property has a corresponding initial weight (usually $w_p = 1$). The distance between two objects can be calculated by various distance metrics. (A variation of the Hamming distance that takes into account the weights of objects has been used for the experiments in the final section of this article.)

The algorithm proceeds in cycles; each cycle comprises of the following sequence of steps:

- Step 1. All the possible threshold values are calculated.
- Step 2. For each of these thresholds, all the pairs of similar objects are determined.
- Step 3. All the maximal sets of entities (categories) that are similar to each other are computed.
- Step 4. For each of the clustering descriptions corresponding to each threshold, a “goodness” value is calculated according to a “goodness” function (see following)
- Step 5. The clustering that rates highest according to the “goodness” function is selected and new weights are calculated according to function (3).
- Step 6. The algorithm is repeated from Step 1 for the new weights.
- Step 7. The algorithm terminates when the newly selected “goodness” value is less than or equal to the value that resulted during the immediately preceding run.

As the *Unscramble* algorithm generates a large number of clusterings (one for each possible similarity threshold), it is necessary to define some measure of “goodness” for each clustering so as to select the best. Two such measures have been considered:

1. **Overlap Function.** This simple function provides a measure for the degree by which clusters overlap; the less overlapping between clusters, the better.

2. **Category Utility.** This function favors categorizations with high uniformity (in terms of properties) within individual clusters (“intra-class similarity”) and strong differences between clusters (“inter-class dissimilarity”). Another way of interpreting this is that category utility measures the *prediction potential* of a categorization: it favors clusterings where it is easy to predict the properties of an entity, given that one knows which cluster it belongs to, and vice versa.

In the experiments reported in the final section, *category utility* has been used. The main advantages of this measure are its firm grounding in statistics, its intuitive semantics, and the fact that it does not depend on any parameters. These measures are discussed in detail in Cambouropoulos & Widmer (2000).

Using the weighting function (Eq. 3), a *prototype* for each cluster can be computed by selecting the most salient properties (values with highest weights). The threshold criterion for each cluster is simply the maximum distance of the prototype to all of the members of the cluster. A new object is a member of a cluster if its distance to the cluster’s prototype is less than or equal to the threshold criterion. If a new object has a new attribute value, this receives a zero initial weight (i.e., it is nondiagnostic). This way *Unscramble* can be used not only for clustering but also for membership prediction tasks.

The most useful characteristics of the *Unscramble* algorithm—depending on the task at hand—are as follows:

- there is no need to define in advance a number of categories
- the prominence of properties is discovered by the algorithm
- categories may overlap
- the description of emerging categories is explicit and can readily be used to predict membership of new objects.

Some examples of the usefulness of such clustering characteristics are presented in the test case given in the next section.

Applying the Clustering Algorithm to a Monophonic Piece by J. S. Bach

In a series of experiments, Deliège (1996, 1997) has studied melodic cue abstraction and categorization processes. For these experiments J. S. Bach’s *Allegro Assai* of the Sonata for Solo Violin in C major BWV 1005 has been used. This monophonic piece is segmented manually into a number of seg-

ments, namely, 26 instances of the two main motives (Figure 4) and 42 one-bar-long motives (for the transitional sections). A more detailed analytic description of this piece is presented in Deliège (1996).

From the various experiments reported in the aforementioned two studies, only two that relate to the current computational experiments are briefly described here (referred to as Experiments 1 and 2 in this paper). In Experiment 1 (Deliège 1996), subjects are asked to listen repeatedly to the two main reference motives (instances 1 and 12 in Figure 4) and then classify the rest of the derivative motives (instances 2–11 and 13–26) into two motive categories (A and B). Musicians perform this task without any mistakes, whereas nonmusicians have a lower success rate.

In Experiment 2 (Deliège 1997), subjects listened only to the first section of this piece (dotted line in Figure 4 separates motive instances of the two sections); then they were presented with items from the first section (Heard), second section (Unheard) and modified items that do not appear in the piece (Modified) and were asked to state whether they had heard these items or not. “Evidence for the formation of an imprint will be considered to be present if, at the end of this incomplete listening, the subjects ‘erroneously recognise’ as having heard excerpts coming from the as-yet-unheard section of the piece” (Deliège 1997, p. 61). In addition, the specific modified items of this experiment should be singled out as being unheard if the hypothesis that “an imprint will reflect the style of the piece” is correct. The results supported these assumptions (the best results, that is, correct responses, were given for the Heard and the Modified items).

The preceding two experiments prompted the design and realization of two similar computational experiments. These are not exact replications of the empirical experiments for reasons that will be explained.

COMPUTATIONAL EXPERIMENT 1

Representing melodic segments is a complex issue (see Cambouropoulos, Crawford, & Iliopoulos, 2001). In the following experiments, motives are represented as lists of musical attributes that reflect melodic and rhythmic aspects of the musical surface. (These representations are kept very simple so that the employment of the algorithm remains as intuitively intelligible as possible.) According to Deliège’s analytic description of the piece, each motive can be subdivided into two 1-bar-long cells, *a* and *b*. The following set of very simple representations was used in this study:

- R1: diatonic pitch, contour, and duration patterns for full-length motive (three attributes)
- R2: diatonic pitch, contour, and duration patterns for full-length motive, for cell *a* and cell *b* (nine attributes)

R3: same as R2, but additionally cells of motive 3 are reversed¹

R4: same as R3 plus 3 extra statistical attributes (*leap*, *rep*, *cdir*) that are meant to reflect melodic properties such as “smoothness,” “repetitiveness,” and “change of direction” (the thresholds below were determined in an ad hoc manner—further research would be necessary to establish a more cognitively pertinent and computationally useful description of such melodic properties):

leap: if number of leaps in motive $\leq 30\%$ then attribute value 1; otherwise value 2

rep: if number of maximally repeated note $\leq 30\%$ then value 1; otherwise value 2

cdir: if number of changes in direction $\leq 50\%$ then value 1; otherwise value 2

Experiment 1 was rather straightforward to replicate on the computer. The distance of each of the motive derivatives (2–11 and 13–26 in Figure 4) to the two main reference motives (1 and 12) was calculated and, then, each derivative was grouped with the reference motive that was closest. The distance metric used is based on the Hamming distance and takes into account the weights of each attribute value (all weights equal to 1 initially in this experiment). For representations R3 and R4, this task was performed correctly for all the instances; for R1 motives 3, 6, and 11, and for R2 motive 3 could not be placed in either category because they have no common attributes (i.e., equal maximum distance) with the two reference motives.

A more complicated and interesting, in computational terms, experiment was also designed. This involved using the *Unscramble* algorithm for clustering all the motives of Figure 4. The *Unscramble* algorithm is used to organize these motives into categories. The algorithm does not know in advance how many clusters it has to construct—it has to determine this automatically. Additionally, it has to determine the most characteristic (diagnostic) attributes and the prototypes for each cluster.

The *Unscramble* algorithm gives the following results for each of the preceding representations (the numbers in each cluster indicate the motives of Figure 4):

R1: {1,2,4,5,7,8,9,10}, {3}, {6,11}, {12,13,14,15,16,17,18,19,20, 21,22,23,24,25,26}

1. A more sophisticated representation would be required to account for “inclusion” relations that are independent of the order of cells. As there is only one instance (motive 3) where such a relation would be useful, a more sophisticated representation was avoided as this would complicate the description of the motives; instead, the two cells of the specific motive were manually reversed.

Motive A

Motive B

f i r s t s e c o n d

1 2 3 4 5 6 7 8 9

12 13 14 15 16 17 18 19 20

s
e
c
t
i
o
n

The image displays a musical score for the first section of the second section of J.S. Bach's *Allegro Assai* for Solo Violin in C major, BWV 1005. The score is presented in two columns. The first column contains measures 10 and 11, which are grouped together by a bracket. The second column contains measures 21, 22, 23, 24, 25, and 26, also grouped together by a bracket. A vertical dotted line is placed between the two columns, separating the first instance of the reference motive (measures 10-11) from its subsequent derivatives (measures 21-26). The notation includes treble clefs, a key signature of one sharp (F#), and various rhythmic values such as eighth and sixteenth notes.

Fig. 4. Motives A and B from J. S. Bach's *Allegro Assai* of the Sonata for Solo Violin in C major BWV 1005. In Deliège's experiment, the first instance of each column is taken to be the reference motive (Motives 1 and 12) and the rest of the instances, their first-degree derivatives. The dotted line separates the motive instances of the first section from the instances of the second section.

- R2: {1,2,4,5,7,8,9,10}, {3}, {6,11}, {12,13,14,15,16,17,18,19,20,21,22,23,24,25,26}
- R3: {1,2,4,5,7,8,9,10}, {1,2,6,7,8,11}, {12,13,14,15,16,17,18,19,20,21,22,23,24,25,26} — first cycle
 {1,2,3,4,5,6,7,8,9,10,11}, {12,13,14,15,16,17,18,19,20,21,22,23,24,25,26} — final
- R4: {1,2,3,4,5,6,7,8,9,10,11}, {12,13,14,15,16,17,18,19,20,21,22,23,24,25,26}

As the representation of the motives in terms of attributes becomes richer, the resulting cluster descriptions improve (they actually give the correct results for R3 and R4). For R1 and R2, motive 3 is a monadic category (this is because its two component cells are reversed) and motives 6 and 11 are clustered together (these two motives end with a quarter-note — actually they end the first and second sections of this piece). It is also clear that the motive B cluster is much “sharper,” that is, the resemblance of its members is much stronger, as it appears even for the simplest representations, R1 and R2.

For each of these clusters, a prototype is computed that comprises the most characteristic attribute values. This prototype may not coincide with any of the existing motives. Of course, anteriority (i.e., the temporal order in which motives appear) is not taken into account, so the reference motives 1 and 12 are not expected to be the prototypes given by the algorithm. For the clustering description produced for R4, the most characteristic attributes are the following:

- Cluster {1,2,3,4,5,6,7,8,9,10,11}: rhythmic pattern of the cell that contains the two eighth-notes, and the three statistical attributes.
- Cluster {12,13,14,15,16,17,18,19,20,21,22,23,24,25,26}: rhythmic pattern of the full-length motive, contour of the first cell and the three statistical attributes.

It can be seen from this experiment that *Unscramble* is successful in clustering the given motives correctly. It finds the correct number of clusters for R3 and R4 and also highlights the most prominent cues that are responsible for clustering the motives together. Some of these cues, such as rhythmic pattern and contour, correspond closely to the cues suggested in the empirical study.

COMPUTATIONAL EXPERIMENT 2

In Experiment 2, described earlier, listeners are merely exposed to the first section of this piece and then asked whether they have heard a number of motivic items presented to them. The task of making sense of a mono-

phonic section by breaking it down into constituent parts is a very complex process at least as far as computational models are concerned. A computational model for melodic segmentation has been developed by Cambouropoulos (1997, 1998b). For practical reasons, in this study, the melodic segments of Section 1 presented in Figure 4 (above the dotted line) are taken as a given (rather than being extracted automatically from the musical surface). These segments (instances 1–6 and 12–18) are presented to *Unscramble* for clustering. Then, a prototype and a threshold criterion are computed for each category. Finally, the Heard items of section 1, the Unheard items of section 2, and the Modified items of Figure 5 are categorized into the existing clusters.²

For this clustering experiment, only representation R4 was used as this was shown to be appropriate in Experiment 1. For this representation, the *Unscramble* algorithm gives the following results:

R4: {1,2,3,4,5,6}, {12,13,14,15,16,17,18}

For each cluster, a prototype is computed and also a threshold criterion is calculated that gives its outermost boundaries.

Then, the distance of each of the Heard, Unheard, and Modified motives to each prototype is calculated; if the distance is less than the threshold criterion, then a motive is said to be a member of that category.

All the Heard motives are classified correctly (trivial), the Unheard motives are classified as expected (i.e., instances 7, 8, 9, 10, and 11 are members of Motive A, and instances 19, 20, 21, 22, 23, 24, 25 and 26 are members of Motive B), and the Modified motives (Figure 5) are not members of either category (as their distance to the prototypes is greater than the threshold criterion).



Fig. 5. Four modified instances of Motive B used in Deliège's experiment (see text).

2. In the empirical study, full-length motives are only one third of the tested motives—and these are considered to be easier to classify than shorter one-bar-long motives.

Even though Experiment 2 is more difficult to replicate and the preceding computational experiment is a gross simplification, it is still very interesting to see that *Unscramble* is capable of inducing descriptions of the emerging categories that can be used successfully both to classify “incorrectly” Unheard instances supporting the hypothesis of imprint formation and to exclude Modified motives that don’t fit in these categories for the particular stylistic context (mainly because of the rhythmic differences).

It would be very interesting to attempt a more sophisticated computational replication of the preceding experiment. Ideally, the system should be able to break down automatically the musical surface into meaningful segments, then construct sophisticated representations for each segment, and finally organize these into motivic categories (Cambouropoulos, 1998a, describes preliminary attempts of this type on other melodic data).

Conclusions

In this article, a theoretical discussion was presented on the various ways the notions of similarity and categorization have been treated by a number of theorists and researchers. Based on this discussion, working formal definitions were given that inextricably bind together similarity and categorization. These definitions readily lend themselves to developing a computational model (the *Unscramble* algorithm) for melodic clustering and membership prediction.

An attempt was made to apply this model to melodic data used in two psychological experiments. Despite the simplifications for the needs of the computational experimentation, it is clear that the results obtained by the application of the proposed algorithm support the underlying hypotheses of the empirical studies on cue abstraction, imprint formation, and categorization (e.g., from various different attributes, *Unscramble* abstracted a number of cues that were appropriate for the specific categorization tasks, organized the given melodic segments into plausible categories, and successfully categorized new melodic material into the previously determined motivic groups). Such computational experiments are interesting because the various stages of the analytic process are transparent to the researcher and the initial hypotheses can be systematically studied.³

3. This research is part of the project Y99-INF, sponsored by the Austrian Federal Ministry of Education, Science, and Culture in the form of a START Research Prize. The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry of Education, Science, and Culture.

References

- Barsalou, L. W. (1992). *Cognitive psychology: An overview for cognitive scientists*. Hillsdale, NJ: Lawrence Erlbaum.
- Beckner, M. (1959). *The biological way of thought*. New York: Columbia University Press.
- Black, M. (1952). The identity of indiscernibles. *Mind*, 61, 153–164.
- Cambouropoulos, E. (1997). Musical rhythm: Inferring accentuation and metrical structure from grouping structure. In M. Leman (Ed.), *Music, gestalt and computing: Studies in systematic and cognitive musicology* (pp. 277–293). Berlin: Springer-Verlag.
- Cambouropoulos, E. (1998a). *Towards a general computational theory of musical structure*. Unpublished doctoral dissertation. University of Edinburgh, United Kingdom.
- Cambouropoulos, E. (1998b). Musical parallelism and melodic segmentation. In A. Argenti & C. Mirolo (Eds.), *Proceedings of the XII Colloquium on Musical Informatics*. Venice, Italy: Associazione di Informatica Musicale Italiana.
- Cambouropoulos, E., Smail, A., & Widmer, G. (1999). A clustering algorithm for melodic analysis. In H. G. Feichtinger & M. Dörfler (Eds.), *Proceedings of the Diderot'99 Forum on Mathematics and Music* (pp. 21–28). Vienna, Austria: Österreichische Computer Gesellschaft.
- Cambouropoulos, E., & Widmer, G. (2000). Automated motivic analysis via melodic clustering. *Journal of New Music Research*, 29(4) (forthcoming).
- Cambouropoulos, E., Crawford, T., & Iliopoulos, C. S. (2001). Pattern processing in melodic sequences: Challenges, caveats and prospects. *Computers and the Humanities*, 35(1) (forthcoming).
- Carterette, E. C., Hohl, D. V., & Pitt, M. A. (1986). Similarities among transformed melodies: The abstraction of invariants. *Music Perception*, 3, 393–410.
- Deliège, I. (1996). Cue abstraction as a component of categorisation processes in music listening. *Psychology of Music*, 24, 131–156.
- Deliège, I. (1997). Similarity in processes of categorisation: Imprint formation as a prototype effect in music listening. In M. Ramscar, U. Hahn, E. Cambouropoulos, & H. Pain (Eds.), *Proceedings of the Interdisciplinary Workshop on Similarity and Categorisation* (pp. 59–65). Edinburgh, U. K.: Department of Artificial Intelligence, University of Edinburgh.
- Deliège, I. (2001). Prototype effects in music listening: An empirical approach to the notion of imprint. *Music Perception*, 18, 371–407.
- Deliège, I., & Mélen, M. (1997). Cue abstraction in the representation of musical form. In I. Deliège & J. Sloboda (Eds.), *Perception and cognition of music* (pp. 387–412). Hove, U.K.: Psychology Press Ltd.
- Eerola, T., Järvinen, T., Louhivuori J, & Toiviainen P. (2001). Statistical features and perceived similarity of folk melodies. *Music Perception*, 18, 275–296.
- Estes, W. K. (1994). *Classification and cognition*. Oxford: Oxford University Press.
- Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the non-independence of features in similarity judgments. *Cognitive Psychology*, 23, 222–262.
- Goodman, N. (1972). Seven strictures on similarity. In *Project and problems* (pp. 437–446). New York: Bobbs-Merrill Co, Inc.
- Hampton, J. (1993). Prototype models of concept representation. In I. Van Mechelen, J. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts* (pp. 67–95). London: Academic Press.
- Hampton, J., & Dubois, D. (1993). Psychological models of concepts: Introduction. In I. Van Mechelen et al. (Eds.), *Categories and concepts* (pp. 11–33). London: Academic Press.
- Hötheker K., Hörnel D., & Anagnostopoulou C. (2001). Investigating the influence of representations and algorithms in music classification. *Computers and the Humanities*, 35(1) (forthcoming).
- Koniari, D., Predazzer S, & Mélen, M. (2001). Categorization and schematization processes used in music perception by 10- to 11-year-old children. *Music Perception*, 18, 297–324.

- Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, 85, 445–463.
- Lakoff, G. (1987). *Women, fire and dangerous things*. Chicago: The University of Chicago Press.
- Lamont, A., & Dibben, N. (1997). Perceived similarity of musical motifs: An exploratory study. In M. Ramscar, U. Hahn, E. Cambouropoulos, & H. Pain (Eds.), *Proceedings of the Interdisciplinary Workshop on Similarity and Categorisation* (pp. 143–149). Edinburgh, U.K.: Department of Artificial Intelligence, University of Edinburgh.
- Lamont, A., & Dibben, N. (2001). Motivic structure and the perception of similarity. *Music Perception*, 18, 245–274.
- Mechelen, I. V., Hampton, J., Michalski, R. S., & Theuns, P. (1993). *Categories and Concepts*. London: Academic Press.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100, 254–278.
- Mélen, M., & Wachsmann, J. (2001). Categorization of musical motifs in infancy. *Music Perception*, 18, 325–346.
- Murphy, G. L. (1993). Theories and concept formation. In I. Van Mechelen, J. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts* (pp. 173–200). London: Academic Press.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289–316.
- Murtagh, F. D. (1993). Cluster analysis using proximities. In I. Van Mechelen, J. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts* (pp. 225–245). London: Academic Press.
- Pollard-Gott, L. (1983). Emergence of thematic concepts in repeated listening to music. *Cognitive Psychology*, 15, 66–94.
- Quine, W. V. (1950). Identity, ostension, and hypostasis. *The Journal of Philosophy*, XLVII, 621–633.
- Rips, L. J. (1989). Similarity, typicality and categorisation. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). Cambridge, U.K.: Cambridge University Press.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192–232.
- Russell, B. (1949). *A critical exposition of the philosophy of Leibniz*. London: Bradford and Dickens.
- Russell, B., & Whitehead, A. N. (1964). *Principia mathematica*. 2nd ed. London: Allen & Unwin.
- Shepard, R. N. (1962a). The analysis of proximities: Multi-dimensional scaling with an unknown distance function, part I. *Psychometrika*, 27, 125–140.
- Shepard, R. N. (1962b). The analysis of proximities: Multi-dimensional scaling with an unknown distance function, part II. *Psychometrika*, 27, 219–246.
- Stroll, A. (1967). Identity. *The encyclopedia of philosophy*. London: Macmillan Publishing Co.
- Sutcliffe, J. P. (1993). Concept, class, and category in the tradition of Aristotle. In I. Van Mechelen, J. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts* (pp. 35–65). London: Academic Press.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352.
- Wittgenstein, L. (1953). *Philosophical investigations*. London: Macmillan Publishing Co.