Diversity and the Geometry of Similarity

Klaus Nehring

Department of Economics, University of California at Davis

Davis, CA 95616, U.S.A.

This Version: July 16, 1999

1 Introduction

In the companion paper "A Theory of Diversity I" (Nehring and Puppe (1999a)), henceforth: TD I, we have studied three paradigmatic classes of diversity functions: hierarchies, lines and the hypercube. They exemplify the primary strategy to endow the general multi-attribute model of diversity with structure while keeping it manageable: single out a subset of all conceivable attributes as relevant by setting the weight of the others to zero. In this paper, we will develop this strategy in generality using the language and tools of abstract convexity theory. The methodology allows one to characterize a variety of naturally patterned families of relevant attributes in terms of conditional independence properties of the corresponding diversity function. The characterization of the line model in terms of the Line Independence condition provided in TD I is in fact an application of this methodology.

Abstract convexity theory supplies also a sound way to construct complex structures from simple ones as their "qualitative product." A product setting offers yet another natural way of imposing structure: independence across dimensions, leading to the definition and characterization of an "independent product" of which the independent hypercube studied in TD I is an instance.

2 Background: Convex Structures Described by Ternary Relations

Our goal is to give a rigorous foundation of a general view of diversity as "overall" dissimilarity. Intuitively, different contexts will be characterized by different patterns of similarity and dissimilarity to be captured by different models. As in TD I, define a dissimilarity (pseudo-)metric d from a given diversity function $v : 2^X \to \mathbf{R}$ by $d(x,y) := v(\{x,y\}) - v(\{y\})$. The notion of a "pattern" of similarity is naturally described by a ternary relation T with the following interpretation. For all objects x, y, z in a given universe X, $(x,y,z) \in T$ if y is more similar to x than z is to x. Hence, in contrast to the *quantitative* notion of dissimilarity expressed by d, the ternary relation T describes a qualitative concept of *comparative* similarity between objects.

The relation T and the quantitative dissimilarity metric $d(\cdot,\cdot)$ jointly determine the "conceptual geometry" of the object space. The analogy to the geometry of physical space may be useful, where d corresponds to distance and T corresponds to betweenness of objects in the sense that $(x,y,z) \in T$ whenever y lies between x and z. While helpful, the analogy is imperfect, e.g. dissimilarity is not symmetric in general.[1]

Example 1 Consider a set $(X, \geq)$ of objects linearly ordered by some characteristic, such as temperature or geographic altitude. Here, qualitative similarity ("line betweenness") is naturally given by

$$(x,y,z) \in T_{\mathcal{L}} :\Leftrightarrow [x \geq y \geq z \text{ or } z \geq y \geq x],$$

i.e. y is more similar than z to x whenever y is between x and z in terms of the characteristic.

Example 2 As another example of how ternary relations can be used to represent

[1] Indeed, as observed in TD I, the quantitative dissimilarity (pseudo-)metric d is symmetric only in the *uniform* case where $v(\{x\}) = v(\{y\})$ for all $x, y \in X$.

qualitative information about objects, consider the following segment of an evolutionary tree (as displayed in the Museum of Natural History, New York City)

Figure 1: A segment of an evolutionary tree

The qualitative similarity information in this tree is described by a ternary relation T_{ev} as follows: $(x, y, z) \in T_{ev}$ if species[2] y branched off later than (or at the same time as) species z from species x. It is easily seen that evolutionary trees are uniquely determined by their induced similarity relation. In the example, one has the following (non-trivial) instances of qualitative similarity: (*salmon, human, shark*), (*shark, human, salmon*), (*shark, salmon, human*) and (*human, salmon, shark*).

In the context of the multi-attribute approach presented in TD I, the conceptual space is described in terms of a family $\mathcal{A} \subseteq 2^X$ of relevant attributes. Any such family of subsets naturally induces a notion of qualitative similarity as follows.

$$(x, y, z) \in T_{\mathcal{A}} :\Leftrightarrow [\text{ for all } A \in \mathcal{A} : \{x, z\} \subseteq A \Rightarrow y \in A]. \tag{2.1}$$

The definition expresses an understanding of similarity as commonality of attributes: For y to be more similar to than z to x, y must possess every attribute shared by x and z. As an example, consider whales (wh) and sharks (sh) in the company of rhinos (rh). Suppose that the only relevant attribute is "being a mammal" corresponding to the case where $\mathcal{A}$ consists only of the subset $\{wh, rh\}$ of all mammals (see TD I, Sect. 2.1, for a discussion of the extensional interpretation of attributes). One has $\{(wh, rh, sh), (rh, wh, sh)\} \subseteq T_{\mathcal{A}}$ and $\{(wh, sh, rh), (rh, sh, wh)\} \cap T_{\mathcal{A}} = \emptyset$, i.e. whales and rhinos are strictly more similar to each other than they are to sharks. Observe that judgements on qualitative similarity will typically change with the inclusion of further attributes. For instance, suppose that in addition to "being a mammal" the attribute $\{wh, sh\}$ ("living in the ocean") is deemed relevant, so that $\mathcal{A}' = \{\{wh, rh\}, \{wh, sh\}\}$. In this case, one obtains $(wh, rh, sh) \notin T_{\mathcal{A}'}$, i.e. rhinos are no longer more similar to whales than sharks are. The example thus also illustrates the general fact that more relevant attributes typically entail fewer qualitative similarity judgements. Of course, this is already apparent from (2.1) since each attribute can be viewed as a "test" that has to be passed by any triple in $T_{\mathcal{A}}$.

In the case of a line, selecting the intervals $\mathcal{L} := \{[x, y] : x, y \in X, y \geq x\}$ as the family of relevant attributes amounts to exactly the direct definition of $T_{\mathcal{L}}$ given in Example 1 above. Also observe that a family $\mathcal{A}$ satisfies the Interval Property IP with respect to the line structure (cf. TD I, Sect. 4) if and only if $T_{\mathcal{A}} \supseteq T_{\mathcal{L}}$.

[2] As in TD I, the term "species" is used as a layman's, not a biologist's notion.

In the context of an evolutionary tree, T_{ev} as specified in Example 2 is derived from a family of attributes $\mathcal{A}_{ev}$ as follows. For any set of species A, let $A \in \mathcal{A}_{ev}$ if and only if A is a singleton or A is the set of all successors of some node (= point of branching). Then, $T_{ev} = T_{\mathcal{A}_{ev}}$. In the above example, $\mathcal{A}_{ev}$ thus consists of all singletons, the universal set, and the set $\{$*salmon, human*$\}$.

Example 3 Consider the hypercube $\{0,1\}^K$. The canonical ternary relation $T_{\mathcal{C}}$ associated with the hypercube ("cube betweenness") is

$$(x, y, z) \in T_{\mathcal{C}} :\Leftrightarrow [\text{ for all } k : x^k = z^k \Rightarrow y^k = x^k = z^k],$$

that is, y is more similar to x than z is if and only if y shares every *property* that is shared by x and z. Clearly, this definition of qualitative similarity in the hypercube corresponds to selecting the family $\mathcal{C}$ of all subcubes as the family of relevant attributes. In particular, a family $\mathcal{A}$ satisfies the Subcube Property SP with respect to the hypercube structure (cf. TD I, Sect. 5) if and only if $T_{\mathcal{A}} \supseteq T_{\mathcal{C}}$.

For any family $\mathcal{A}$, the ternary relation $T_{\mathcal{A}}$ satisfies the following three properties.

T1 (Reflexivity) $y \in \{x, z\} \Rightarrow (x, y, z) \in T$.
T2 (Symmetry) $(x, y, z) \in T \Leftrightarrow (z, y, x) \in T$.
T3 (Transitivity) $[(x, x', z) \in T$ and $(x, z', z) \in T$ and $(x', y, z') \in T] \Rightarrow (x, y, z) \in T$.

Properties T1 and T2 follow immediately from the definition of $T_{\mathcal{A}}$. T1 is largely a matter of convention. The most forceful condition is perhaps T2 which justifies the geometric interpretation of T as betweenness relation.[3] The ternary relation induced by the line structure in Example 1 is in fact *the* classic instance of a betweenness relation. This example also illustrates well the intuitive content of property T3: If both x' and z' are between x and z, and moreover y is between x' and z', then y must also lie between x and z. The validity of T3 in general can be verified without much difficulty.

In the following we will refer to ternary relations satisfying T1, T2 and T3 as *ternary similarity orderings* (TSOs). One has the following result which we state without proof (cf. Nehring (1997)).

Fact 2.1 *$T = T_{\mathcal{A}}$ for some attribute family $\mathcal{A}$ if and only if T is a TSO.*

For given T and x, denote by T^x the induced binary similarity relation with respect to x, i.e. $yT^xz :\Leftrightarrow (x, y, z) \in T$. It is easily verified that for any TSO T and all $x \in X$, T^x is reflexive and transitive. On the other hand, binary similarity comparisons according to T^x are typically incomplete, since often y and z will share attributes with x that are not shared with each other. For instance with "being a mammal" and "ocean-living" as relevant attributes, neither are sharks more similar than rhinos to whales, nor are rhinos more similar than sharks to whales. The following result shows that completeness of T^x for all x in fact characterizes the class of all hierarchies.[4]

Proposition 2.1 *A family $\mathcal{A}$ of subsets of X is a hierarchy if and only if, for all $x \in X$, $T^x_{\mathcal{A}}$ is complete.*

[3] In the literature, symmetric ternary relations are often simply called "betweenness relations." Betweenness relations in this sense have been introduced into the axiomatic foundations of geometry by Pasch [1882] and frequently employed since then (see, e.g. Hilbert [1899], Suppes [1972], Fishburn [1985, ch.4]).

[4] Recall that $\mathcal{A}$ is a *hierarchy* if and only if, for all $A, B \in \mathcal{A}$, $A \cap B \neq \emptyset \Rightarrow [A \subseteq B$ or $B \subseteq A]$.

So far, we have studied the properties of the qualitative similarity relation $T_{\mathcal{A}}$ induced by a given family of relevant attributes. Our purpose is to characterize different diversity models in terms of the associated TSO. To this end one also needs to consider the converse problem: Which attributes are compatible with a given TSO? Formally, an attribute A will be called *compatible* with the ternary relation T if for all x, y, z,

$$(x, y, z) \in T \Rightarrow [\{x, z\} \subseteq A \Rightarrow y \in A]. \tag{2.2}$$

Under a betweenness interpretation of T, compatibility in this sense can be viewed as "convexity" with respect to T: a set A is compatible with T if, for any $x, z \in A$, the set A also contains any point that is T-between x and z. For given T, let $\mathcal{A}_T$ denote the family of all $A \in 2^X$ that are compatible with T. For instance, the families of all subsets compatible with the line betweenness $T_{\mathcal{L}}$ and the cube betweenness $T_{\mathcal{C}}$ are the family $\mathcal{L}$ of all intervals and the family $\mathcal{C}$ of all subcubes, respectively. Any family $\mathcal{A}_T$ derived from some ternary relation T satisfies the following three properties.

A1 (Boundedness) $\mathcal{A} \supseteq \{\emptyset, X\}$.
A2 (Closedness under Intersections) $A, B \in \mathcal{A}$ implies $A \cap B \in \mathcal{A}$.
A3 (2-arity) $A \in \mathcal{A}$ whenever, for all $x, y \in A$, there exists $B \in \mathcal{A}$ such that $\{x, y\} \subseteq B \subseteq A$.

In a finite setting, families of sets satisfying A1, A2 are commonly called "convex structures" in the literature on abstract convexity (see, e.g. van de Vel (1993)). Those satisfying A3 in addition are called "2-ary" convex structures. For brevity, we omit explicit reference to "2-arity" speaking of the latter simply as *convex structures* (CVS). Note that, due to A1 and A2, any convex structure forms a lattice with intersection defining the meet and in which the join of two attributes is defined as the smallest common super-attribute. A3 can be paraphrased as saying that a set A is "convex" (i.e. belongs to $\mathcal{A}$) whenever, for any two of its elements, A contains their join which can be viewed as the "segment" spanned by them.

Intersection-closedness A2 is a natural property of classes of relevant attributes, as it corresponds to conjunction-closedness of the attribute-defining features; for example, if "being a mammal" and "living in the ocean" are features deemed relevant, so will be presumably the feature "being an ocean-living mammal." This also explains the intuitive "uniqueness" of objects as due to their possessing a unique *combination* of features, while they may share any particular feature with many objects. By contrast, closedness under union (or under complementation) is not desirable as a general property as it would generate many artificial attributes such as "being a mammal *or* living in the ocean." The philosopher Gärdenfors has argued in a series of papers in a related vein (see, e.g. Gärdenfors (1990)) that legitimate inductive inference needs to be based on convex predicates.[5]

To illustrate the role of condition A3 (2-arity), consider the family $\mathcal{A}$ of all line segments in $\mathbf{R}^K$. The induced TSO $T_{\mathbf{R}^K} := T_{\mathcal{A}}$ according to (2.1) is given by the standard notion of Euclidean betweenness in $\mathbf{R}^K$: $(x, y, z) \in T_{\mathbf{R}^K} \Leftrightarrow y = \beta x + (1 - \beta) z$ for some $\beta \in [0, 1]$. The family of all subsets that are compatible with $T_{\mathbf{R}^K}$ in the sense of (2.2) is the set of all convex sets in $\mathbf{R}^K$ in the usual sense. In particular, the set of all convex sets in $\mathbf{R}^K$ is the smallest *2-ary* convexity that contains all line segments.

[5] For readers familiar with the relevant philosophical literature, Goodman's provocative predicate "grue" (= green before date t, blue after t), for instance, is construed by Gärdenfors as non-convex. Similarly, in Hempel's paradox the predicate "non-raven" is identified as non-convex, but "non-black" as convex.

For a proof of the following two results, see Nehring (1997).

Fact 2.2 *$\mathcal{A} = \mathcal{A}_T$ for some ternary relation T if and only if $\mathcal{A}$ is a CVS.*

TSOs and CVSs are related by an order-inverting isomorphism:

Fact 2.3 *The mapping $\mathcal{A}_\bullet : T \mapsto \mathcal{A}_T$ is an order-inverting bijection*[6] *between TSOs and CVSs whose inverse is given by $T_\bullet : \mathcal{A} \mapsto T_\mathcal{A}$. In particular, $T_{\mathcal{A}_T} = T$ for any TSO T, and $\mathcal{A}_{T_\mathcal{A}} = \mathcal{A}$ for any CVS $\mathcal{A}$.*

3 The Geometry of Similarity

3.1 Conditional Independence as the Qualitative Structure of Diversity Functions

What is the qualitative structure of a diversity function $v : 2^X \to \mathbf{R}$? An indirect answer to this suggests itself in the form of the induced family of relevant attributes, i.e. the support Λ of its conjugate Moebius inverse λ. This has been one major theme of TD I.[7] We will now show that one can formally describe the qualitative structure of a diversity function in terms of properties of the function itself. A characteristic property of diversity functions is their submodularity, and significant qualtitative information is obtained by determining where submodularity is degenerate. Recall that, in terms of the distinctiveness function $d : X \times 2^X \to \mathbf{R}$, defined as

$$d(x, S) := v(S \cup \{x\}) - v(S) = \lambda(\{A : x \in A, A \cap S = \emptyset\},$$

submodularity is the requirement that, for all x, $d(x, S)$ is non-increasing in its second argument. In particular, distinctiveness of x from S cannot increase with the inclusion of z: for all x, z, S, $d(x, S) \geq d(x, S \cup \{z\})$. Submodularity is degenerate at x, z, S if $d(x, S) = d(x, S \cup \{z\})$, i.e. if the distinctiveness of x from S is not diminished by the inclusion of z. One has,

$$\begin{aligned} &d(x, S) - d(x, S \cup \{z\}) \\ &\qquad = \lambda(\{A : x \in A, A \subseteq S^c\}) - \lambda(\{A : x \in A, A \subseteq S^c, z \notin A\}) \\ &\qquad = \lambda(\{A : \{x, z\} \subseteq A \subseteq S^c\}), \end{aligned}$$

where S^c denotes the complement of S in X. Hence, by non-negativity of λ, strict submodularity at x, z, S results from the existence of relevant attributes that *jointly* distinguish x and z from S; equivalently,

$$d(x, S) = d(x, S \cup \{z\}) \Leftrightarrow [\text{ for no } A \in \Lambda : \{x, z\} \subseteq A \subseteq S^c]. \tag{3.1}$$

This motivates the following definition. Let v be a diversity function on 2^X. We say that x is *independent from z conditional on (the inclusion of) y*, denoted by $(x, y, z) \in T_v$,

[6] i.e. $T' \subseteq T \Rightarrow \mathcal{A}_{T'} \supseteq \mathcal{A}_T$, and conversely, $\mathcal{A}' \subseteq \mathcal{A} \Rightarrow T_{\mathcal{A}'} \supseteq T_\mathcal{A}$.

[7] Recall that, by definition, $v : 2^X \to \mathbf{R}$ is a *diversity function* if there exists a non-negative measure λ on 2^X, referred to as the *conjugate Moebius inverse*, such that, for all S, $v(S) = \lambda(\{A : A \cap S \neq \emptyset\})$. The support $\Lambda := \{A : \lambda_A \neq 0\}$ is referred to as the corresponding *family of relevant attributes.*

if the distinctiveness of x from any set S that includes y does not change with the addition of z to S. Formally,

$$(x, y, z) \in T_v :\Leftrightarrow \text{ for all } S \ni y, d(x, S) = d(x, S \cup \{z\}). \tag{3.2}$$

Observe that $(x, y, z) \in T_v \Leftrightarrow d(x, y) = d(x, \{y, z\})$, hence by (3.1),

$$(x, y, z) \in T_v \Leftrightarrow [\text{ for all } A \in \Lambda : \{x, z\} \subseteq A \Rightarrow y \in A],$$

i.e. $T_v = T_\Lambda$. Summarizing, we have established the following result.

Theorem 3.1 *For any diversity function* $v : 2^X \to \mathbf{R}$, $T_v = T_\Lambda$.

By Theorem 3.1, the "geometry" of the family of relevant attributes $\Lambda = \text{supp}\lambda$ described by T_Λ according to (2.1) is mirrored in the conditional independence relation T_v derived from v. In particular, by Fact 2.1, for any diversity function v, T_v is a TSO.[8]

3.2 Adapting a Model to a Geometry

Suppose a modeller accepts a certain geometric description T of qualitative similarity between objects. What restrictions on her diversity assessments does this entail? Say that a diversity function v is *compatible* with T if $T_v \supseteq T$, i.e. x is independent from z conditional on the inclusion of y whenever y is more similar than z to x. This makes sense since, in view of Theorem 3.1, T_v can be read as the qualitative similarity implicit in v. Compatibility in this sense means that any similarity given by T is respected by v. For instance, in the case of a line with $T_\mathcal{L}$ defined as in Example 1 above, compatibility of a diversity function v translates into the following condition,

$$y \in [x, z] \Rightarrow \text{ for all } S \ni y, d(x, S) = d(x, S \cup \{z\}).$$

Noting that the equality on the right hand side can be rewritten as

$$v(S \cup \{x\}) + v(S \cup \{z\}) = v(S) + v(S \cup \{x, z\}),$$

this condition is easily seen to be equivalent to the "Line Equation"

$$v(\{x_1, ..., x_m\}) = v(\{x_1\}) + \sum_{i=2}^{m} d(x_i, x_{i-1}),$$

where $x_1 < x_2 < ... < x_m$ (cf. TD I (4.1)).

Compatibility in the sense that $T_v \supseteq T$ makes essential use of the fact that the function $v : 2^X \to \mathbf{R}$ is cardinally scaled. In a decision making context, this is justified by viewing v as a von-Neumann-Morgenstern utility function as in TD I. In such contexts, one can go further and define conditional independence and thus compatibility in terms of the ultimate primitive, the decision maker's preference relation over set-lotteries (cf. TD I, Sect. 2.4) as follows. Define a ternary relation $T_\succeq$ by

$$(x, y, z) \in T_\succeq :\Leftrightarrow \text{ for all } S \ni y : \left[\frac{1}{2} \cdot \mathbf{1}_{S \cup \{x\}} + \frac{1}{2} \cdot \mathbf{1}_{S \cup \{z\}}\right] \sim \left[\frac{1}{2} \cdot \mathbf{1}_S + \frac{1}{2} \cdot \mathbf{1}_{S \cup \{x, z\}}\right].$$

From the above discussion, one immediately obtains the following result.

[8] In fact, it can be shown that Theorem 3.1 generalizes to arbitrary set functions, see Nehring and Puppe (1999b).

Fact 3.1 *For any von-Neumann-Morgenstern utility function v corresponding to $\succeq$, $T_{\succeq} = T_v$.*

Observe that the condition LI (Line Independence) in TD I, Sect. 4, is nothing but the statement "$T_{\succeq} \supseteq T_{\mathcal{L}}$."

The restriction on the family of relevant attributes induced by compatibility of v with a certain geometry are characterized by the following result.

Theorem 3.2 *Let T be a TSO. Then v is compatible with T, i.e. $T_v \supseteq T$, if and only if $\Lambda \subseteq \mathcal{A}_T$.*

In the line example, with $T = T_{\mathcal{L}}$ as defined above, the equivalence of the Interval Property "$\Lambda \subseteq \mathcal{L}$" and the Line Independence condition "$T_{\succeq} \supseteq T_{\mathcal{L}}$" (cf. TD I, Th. 4.1) thus follows as a corollary from Theorem 3.2 using Fact 3.1.

The characterization of the hypercube in terms of risk-neutrality properties of the underlying preferences follows in an analogous way: By Theorem 3.2 and Fact 3.1, the Subcube Property $\Lambda \subseteq \mathcal{C}$ is equivalent to $T_{\succeq} \supseteq T_{\mathcal{C}}$ ("Cube Independence").

Returning to Example 2 of an evolutionary tree, one obtains from Theorem 3.2 that $\Lambda \subseteq \{\{$*shark, salmon, human*$\}, \{$*salmon, human*$\}, \{$*shark*$\}, \{$*salmon*$\}, \{$*salmon*$\}\}$ for any diversity function that is compatible with T_{ev}. Thus, given uniformity (i.e. equal valuation of singletons), accepting T_{ev} as the qualitative similarity relation entails that the set $\{$*salmon, human*$\}$ is (weakly) less diverse than the set $\{$*shark, salmon*$\}$. The decision maker may feel this to be inappropriate, which shows that, on reflection, she cannot really accept T_{ev} as the "right" description of qualitative similarity. The obvious resolution is that there has to be an attribute jointly shared by sharks and salmons ("being a fish") that is not shared by humans. Including such attribute removes qualitative similarities from T_{ev}.[9]

In view of Fact 2.3, Theorem 3.2 can be restated as follows.

Theorem 3.2′ *Let $\mathcal{A}$ be a CVS. Then $\Lambda \subseteq \mathcal{A}$ if and only if v is compatible with $T_{\mathcal{A}}$, i.e. $T_v \supseteq T_{\mathcal{A}}$.*

Theorem 3.2 starts with a geometric description given by some TSO T and, using conditional independence, arrives at restrictions on the set of relevant attributes. Theorem 3.2′, in contrast, starts from such restrictions and specifies the TSO needed to obtain them via conditional independence.

To illustrate, consider a hierarchical family $\mathcal{H}$ of attributes and a diversity function v. By Theorem 3.2′, $\Lambda \subseteq \mathcal{H}$ implies $T_v \supseteq T_{\mathcal{H}}$. The latter condition is easily seen to imply the recursion formula $d(x, S) = \min_{y \in S} d(x, y)$ (cf. TD I, Sect. 3): Let $y^* \in S$ be the "$T_{\mathcal{H}}$-nearest" element of S to x, i.e. $y^* T^x_{\mathcal{H}} y$ for all $y \in S$. Such y^* exists by completeness of $T^x_{\mathcal{H}}$ in the hierarchical case (cf. Proposition 2.1 above). Since $T_v \supseteq T_{\mathcal{H}}$, one obtains $d(x, W) = d(x, W \cup \{y\})$ for all $W \ni y^*$ and all $y \in S$. Hence, by induction, $d(x, y^*) = d(x, S)$. By submodularity, $d(x, y^*) = \min_{y \in S} d(x, y)$.

3.3 Attribute Structure Revealed

By Theorems 3.2 and 3.2′, the conditional independence relation T_v entails information on the family Λ of relevant attributes, but constrains Λ only "approximately." The

[9] Observe, that including biological taxa in the set of relevant attributes may destroy the hierarchical structure of the evolutionary tree model. Given certain compatibility assumptions, the resulting structure may nevertheless be compatible with the line model.

reason is that different families Λ may induce the same qualitative similarity relation T_Λ. As an example, consider the three families: $\Lambda_1 = 2^X$, $\Lambda_2 = \{X \setminus \{x\} : x \in X\}$ and $\Lambda_3 = \{\{x, y\} : x, y \in X\}$. As is easily verified, one has $T_{\Lambda_1} = T_{\Lambda_2} = T_{\Lambda_3} = T^\emptyset$, where $T^\emptyset := \{(x, y, z) : y \in \{x, z\}\}$ is the trivial similarity relation that only accounts for the reflexivity condition T1.

The extent to which the family of relevant attributes is determined by the associated conditional independence relation can be made precise. Specifically, it will be shown that T_v reveals the support Λ of the conjugate Moebius inverse up to "abstract convexification." To define the latter, one needs to introduce the notion of *CVS-closure* of a family of attributes. For any $\mathcal{A} \subseteq 2^X$, let $\mathcal{A}^*$ denote the smallest CVS containing $\mathcal{A}$. This is well-defined due to the following fact whose verification is straightforward.

Fact 3.2 *The class of CVSs is closed under intersection.*

The following is the key observation of this subsection.

Theorem 3.3 *For any diversity function v, $\Lambda^* = \mathcal{A}_{T_v}$.*

To illustrate the content of Theorem 3.3, consider a given hypercube and assume that for a diversity function v satisfying the Subcube Property SP one has $T_v = T_\mathcal{C}$, i.e. T_v coincides with the cube betweenness (cf. Example 3 above). It can be verified that in this case, the corresponding Λ must contain all half-spaces. By intersection-closedness, this implies $\Lambda^* = \mathcal{C} = \mathcal{A}_{T_v}$, as asserted by the theorem.

While the hypercube illustrates the typical case in which there is a range of indeterminacy of the support, in the special case of a hierarchy the independence relation T_v fully reveals the support (up to the inclusion of $\emptyset$ and X). In fact, this property characterizes the hierarchy.

Proposition 3.1 *A family $\mathcal{A} \supseteq \{\emptyset, X\}$ has the property that, for any family $\mathcal{A}' \supseteq \{\emptyset, X\}$, $[T_{\mathcal{A}'} = T_\mathcal{A} \Rightarrow \mathcal{A}' = \mathcal{A}]$ if and only if $\mathcal{A}$ is a hierarchy.*

3.4 Qualitative and Quantitative Similarity

Intuitively, the relation T_v $(= T_\Lambda)$ induced by a diversity function v can be viewed as the "qualitative core" of the corresponding quantitative dissimilarity metric d. The following result is an immediate consequence of submodularity of a diversity fuction.

Fact 3.3 *The dissimilarity metric associated with v is* adapted *to T_v in the sense that*

$$(x, y, z) \in T_v \Rightarrow d(x, y) \leq d(x, z). \tag{3.3}$$

Hence, greater qualitative dissimilarity implies greater quantitative dissimilarity.

In the context of "physical" geometry, the canonical way to obtain a ternary betweenness relation from a given distance function is as *geodesic betweenness*: y is geodesically between x and z if $d(x, z) = d(x, y) + d(y, z)$, which has the interpretation of y lying on a shortest path from x to z (see Menger (1928)). In the context of similarity, i.e. conceptual rather than physical geometry, additivity holds only exceptionally, as we have argued in TD I (Sect. 4.3). A natural definition of betweenness T_d induced by a dissimilarity metric is as follows,

$$(x, y, z) \in T_d :\Leftrightarrow [d(x, y) \leq d(x, z) \text{ and } d(z, y) \leq d(z, x)].$$

Thus, y is *metrically between* x and z if, quantitatively, y is both less dissimilar than z to x, and less dissimilar than x to z. In the definition of T_d we view d as derived from a function v that is defined on a domain containing the family $\mathcal{B} := \{\{x, y\} : x, y \in X\}$ of all subsets with at most two elements.

It is obvious that d is adapted to T_d in the sense of (3.3); in fact, T_d is the largest symmetric ternary relation to which d is adapted. While reflexive and symmetric, for general d (derived from some v), T_d need not be transitive, and hence is not necessarily a TSO. Note further that T_d can be equivalently described in terms of *quantitative similarity* $\sigma(x, y) := v(\{x\}) - d(x, y)$ (cf. TD I, Sect. 2.2):

$$(x, y, z) \in T_d \Leftrightarrow \sigma(x, z) \leq \min\{\sigma(x, y), \sigma(y, z)\}. \tag{3.4}$$

The equivalence follows at once from the observation that, in contrast to d, the similarity function σ is always symmetric. In particular, (3.4) shows that T_d is *triple-connected* in the sense that at least one of the three triples (x, y, z), (y, z, x), (z, x, y) is in T_d.

Since T_v is a TSO, one always has $T_v \subseteq T_d$. Indeed, $(x, y, z) \in T_v$ implies $(z, y, x) \in T_v$ by symmetry (T2), hence $(x, y, z) \in T_d$ by Fact 3.3. Since T_d is triple-connected, a necessary condition for T_d to coincide with T_v is that T_v is triple-connected as well. The following result shows that this condition is also sufficient.

Proposition 3.2 *Let $v : 2^X \rightarrow \mathbf{R}$ be a diversity function; then $T_d = T_v$ if and only if T_v is triple-connected.*

Hence, if T_v is triple-connected, T_d is the canonical notion of betweenness described by T_v ($= T_\Lambda$); in particular, T_d is a TSO in that case. The hypercube as an instance of the general case in which T_v is not triple-connected, and hence a proper subrelation of T_d. On the other hand, it is easily verified that T_v derived from a diversity function on a line is triple-connected. One thus obtains the following corollary.

Corollary 3.1 *Let $v : 2^X \rightarrow \mathbf{R}$ be a diversity function satisfying the Interval Property ($T_v \supseteq T_{\mathcal{L}}$) with respect to some linear order $\geq$ on X. Then, $T_d = T_v$.*

3.5 Dissimilarity Metrics Consistent with the Line and Hierarchy Models

In Section 4 of TD I, we have characterized the restrictions on a dissimilarity metric for it to be consistent with a *given* line model associated with a particular linear ordering of the object space. A more fundamental question would address the restrictions of the line model as such: Under what conditions on a dissimilarity metric d is there *some* linear ordering $\geq$ of the object space such that d is the dissimilarity associated with a diversity function v that is compatible with the betweenness induced by $\geq$? This is a complex problem, and we provide an answer for the two polar and most interesting cases of hierarchies and "exact" lines, i.e. the case where *every* interval corresponds to a relevant attribute ($\Lambda = \mathcal{L}$). We consider the case of hierarchies first.

Theorem 3.4 *A function $v^{\mathcal{B}} : \mathcal{B} \rightarrow \mathbf{R}$ can be uniquely extended to a hierarchical diversity function on 2^X if and only if the induced dissimilarity metric d is non-negative, bounded (in the sense that $d(x, y) \leq v(\{x\})$) and, for all x, T_d^x is complete.*

Theorem 3.4 generalizes a classical result (Johnson (1967), Benzecri *et al.* (1973)) on the representation of ultrametric distance functions by not assuming symmetry of d. A symmetric distance function d is called *ultrametric* if, for all x, y, z,

$$\text{mid}\{d(x,y), d(y,z), d(x,z)\} = \max\{d(x,y), d(y,z), d(x,z)\}, \tag{3.5}$$

i.e. if the two greatest distances between any three points are equal. The notion of ultrametricity can be generalized to the case where d is not necessarily symmetric. Say that the quantitative similarity function σ associated with $v^{\mathcal{B}}$ is *ultrametric* if

$$\text{mid}\{\sigma(x,y), \sigma(y,z), \sigma(z,y)\} = \min\{\sigma(x,y), \sigma(y,z), \sigma(z,y)\}. \tag{3.6}$$

In the symmetric case, one has $v(\{x\}) = v(\{y\})$ for all x, y, hence (3.6) and (3.5) are equivalent. It is also easily verified that, in the general case, (3.6) is equivalent to completeness of T_d^x, for all x.

For the polar case of an exact line one has the following result. Say that a ternary relation T is *line-transitive* if, for all x, y, z, w with $y \neq z$,

$$[(x,y,z) \in T \text{ and } (y,z,w) \in T] \Rightarrow (x,y,w) \in T.$$

Furthermore, a dissimilarity metric d is *line-submodular* with respect to T if the following condition holds (cf. TD I, Sect. 4.4). For all x_1, x_2, x_3, x_4 such that $(x_i, x_j, x_l) \in T$ whenever $1 \leq i < j < l \leq 4$,

$$d(x_1, x_4) - d(x_1, x_3) \leq d(x_2, x_4) - d(x_2, x_3).$$

Theorem 3.5 *A function $v^{\mathcal{B}} : \mathcal{B} \to \mathbf{R}$ can be uniquely extended to a diversity function v on 2^X such that $T_v = T_{\mathcal{L}}$ for $\mathcal{L}$ associated with some linear order $\geq$ on X if and only if T_d is line-transitive, and d is bounded, strictly positive (i.e. $d(x,y) > 0$ whenever $x \neq y$), and line-submodular with respect to T_d.*

The condition that drives the result is line-transitivity. Line-transitivity as a condition on T_v is already quite restrictive, for instance, it is satisfied neither in the hypercube, nor in hierarchies. Line-transitivity is even more powerful when applied to the less regular T_d.

In order to prove Theorem 3.5 one wishes to make use of Theorem 4.3 of TD I, the Line Extension Theorem for a *given* line structure. In view of Corollary 3.1 it remains to find conditions on a ternary relation that are necessary and sufficient for its being the betweenness of a linear ordering. These can be found in the literature (see Krantz, Luce, Suppes and Tversky (1979), Suppes (1972)). The additional contribution of Theorem 3.5 and its proof consists in obtaining several of these conditions from the definition of T_d (in particular, triple-connectedness and transitivity of all T_d^x). The insight of Theorem 3.5 is thus the appropriateness of T_d as the "right" notion of betweenness derived from d in the context of the line structure.

4 Constructing Complex Structures: Taking Products

In many applications the specified models presented so far will be too restrictive; while very well behaved and tractable, the line and hierarchy models cannot adequately describe complex situations. However, they will be suitable to describe certain aspects, or dimensions. Therefore, one needs tools for constructing complex structures from simple ones. If objects can be described in terms of qualities in several dimensions as vectors of characteristics, a natural way to achieve this is by defining an appropriate product operation. An example of product is the hypercube which has the simplest possible constituent structure, namely a binary distinction along each dimension.

4.1 Qualitative Products

4.1.1 Weak Product

Suppose that $X = \prod_{k \in K} X^k$, and let, for each coordinate k, T^k be a given qualitative similarity relation on X^k. A minimal notion of a product of the T^ks is the *weak product* $\times_{k \in K} T^k$, defined as follows. For all $x, y, z \in X$,

$$(x, y, z) \in \times_{k \in K} T^k :\Leftrightarrow [\text{ there exists } k : (x^k, y^k, z^k) \in T^k \text{ and } x^{-k} = y^{-k} = z^{-k}],$$

where x^k denotes the k-th coordinate of x, and x^{-k} is the vector with the k-th coordinate deleted. The weak product thus declares y as more similar than z to x if and only if these three objects differ only in one coordinate and y is more similar than z to x in that coordinate. It is easily verified that the weak product, which we also denote by $T^{\text{weak}} := \times_{k \in K} T^k$, is a TSO whenever all T^k are. A set $A \subseteq X$ is compatible with the weak product ($A \in \mathcal{A}_{T^{\text{weak}}}$) if and only if every one-dimensional section of A is compatible with the corresponding T^k, i.e. if and only if, for every k and every fixed z^{-k}, $A \cap (X^k \times \{z^{-k}\}) \in \mathcal{A}_{T^k}$. As an example consider the weak product of two lines, as depicted in Figure 2.

Figure 2: Weak product of two lines

More concretely, think of the coordinates as representing gender, $g \in \{$*female, male*$\}$, and age, $t \in \mathbf{R}_+$. Denote by $T^\emptyset$ the trivial TSO (i.e. $(x, y, z) \in T^\emptyset \Leftrightarrow y \in \{x, z\}$), and consider $T^{\text{weak}} = T^\emptyset \times T_{\mathcal{L}}$. The readership of "Gone with the Wind" (GwW) presumably depends on gender and age interactively; an empirically plausible specification of its extension is the following, $A_{GwW} = \{(g, t) : t \geq t_g\}$ with $t_{\mathit{female}} < t_{\mathit{male}}$. The assumption is thus that females enjoy "Gone with the Wind" at an earlier age than males. Obviously, A_{GwW} is compatible with T^{weak}.

4.1.2 Separable Product

As illustrated by the above examples the weak product allows for interaction of characteristics. A stronger notion of product would rule this out. To express this idea of separability across dimensions, assume away for the moment any similarity information along coordinates. The absence of interaction between dimensions means that all attributes have to be rectangles, i.e. elements of

$$\mathcal{A}^{\text{sep}} := \{A : A = \prod_k A^k \text{ for some } A^k \subseteq X^k\}.$$

A family $\mathcal{A} \subseteq 2^X$ will be called *separable* if $\mathcal{A} \subseteq \mathcal{A}^{\text{sep}}$. For instance, separability of the set A_{GwW} in the above example would require that $t_{female} = t_{male}$, i.e. that females and males enjoy "Gone with the Wind" from the same age on. One might argue that separability is tautologically applicable provided that characteristics are exhaustively specified. For instance, the interaction between dimensions in the example is due to a relevant characteristic that has been omitted in the description ("latent sentimentality").

It follows from Fact 4.1 below that $\mathcal{A}^{\text{sep}}$ is a CVS. The corresponding TSO $T^{\text{sep}} = T_{\mathcal{A}^{\text{sep}}}$ is given by

$$(x, y, z) \in T^{\text{sep}} \Leftrightarrow [\text{ for all } k : y^k \in \{x^k, z^k\}].$$

Note that in the case of a hypercube, T^{sep} coincides with the cube betweenness $T_{\mathcal{C}}$ defined above (Sect. 2, Example 3).

A set function $v : 2^X \to \mathbf{R}$ will be called separable whenever the corresponding family Λ of relevant attributes is separable. In view of Theorem 3.2, the suitability of the notion of separability in applications can be tested through the conditional independence properties of v induced by T^{sep}. Generalizing the hypercube example, these are given by requiring $d(x, S) = d(x, S \cup \{z\})$, for all $S \ni y$, whenever, for all $k \in K$, $x^k = z^k \Rightarrow y^k = x^k = z^k$. By Theorem 3.2 the class of diversity functions satisfying this is precisely the class of all separable ones.

As an example in the biodiversity context, consider $X = X^1 \times X^2$, where X^1 is a set of species and X^2 a set of "habitats";[10] thus, $x = (x^1, x^2) \in X$ is interpreted as "species in habitat." Specifically, assume that X^1 consists of the species gorilla (go) and chimpanzee (ch), and that $X^2 = \{1, 2\}$ with 1 standing for "fenced" and 2 for "wild." Consider, for instance, $x = (go, 1)$ ("fenced gorilla"), $y = (go, 2)$ ("wild gorilla") and $z = (ch, 2)$ ("wild chimpanzee"). Since separability clearly entails $(x, y, z) \in T^{\text{sep}}$, the associated conditional independence conditions thus require that $d(x, \{y\}) = d(x, \{y, z\})$, i.e. the distinctiveness of "fenced gorilla" from {"wild gorilla"} does not change with the addition of "wild chimpanzee," which seems reasonable.

The combination of this notion of separability with coordinatewise similarity information is achieved as follows. For any ternary relation T, let T^* denote the smallest TSO containing T.[11] The *separable product*, henceforth simply: *product*, $\otimes_{k \in K} T^k$ of a set of TSOs T^k on X^k is defined as the smallest TSO containing both $T^{\text{weak}} = \times_{k \in K} T^k$ and T^{sep}, formally,

$$\otimes_{k \in K} T^k := (T^{\text{weak}} \cup T^{\text{sep}})^*.$$

[10] "Habitat" can of course be given a variety of other interpretations, besides the one in the text. In particular, the interpretations "geographic location" and "ecological habitat" seem to be of significant bio-economical interest; note that in both of these, the universe of habitats is fairly naturally endowed with a convex structure.

[11] This is well defined since, obviously, TSOs are closed under intersection.

Proposition 4.1 *For all* x, y, z,

$$(x, y, z) \in \otimes_{k \in K} T^k \Leftrightarrow [\text{ for all } k : (x^k, y^k, z^k) \in T^k].$$

The proof in the appendix constructs $\otimes_k T^k$ from T^{weak} and T^{sep} using the transitivity condition T3.

Fact 4.1 *For* $k \in K$*, let* T^k *be a qualitative similarity relation on* X^k*. Denote by* $\mathcal{A}^k$ *the corresponding CVS, i.e.* $\mathcal{A}^k := \mathcal{A}_{T^k}$*. Then,*

$$\mathcal{A}_{(\otimes_{k \in K} T^k)} = \otimes_{k \in K} \mathcal{A}^k := \{A : A = \prod_{k \in K} A^k, A^k \in \mathcal{A}^k\}.$$

Note that $T^{\text{sep}} = T^{\emptyset} \otimes ... \otimes T^{\emptyset}$, hence it follows from Fact 4.1 that $\mathcal{A}^{\text{sep}}$ is a CVS as claimed above. Also observe that, by definition of $\otimes_k T^k$ and the fact that T_v is a TSO (cf. Th. 3.1 above), compatibility of v with the product similarity is equivalent to compatibility of v with the weak product plus separability.

Proposition 4.1 and Fact 4.1 together show that the canonical notion of a product in abstract convexity theory $\otimes_{k \in K} \mathcal{A}^k$ combines the purely mathematical notion of "embedding" captured by the weak product with the substantive notion of "separability across dimensions" captured by the separable convexity $\mathcal{A}^{\text{sep}}$.

In the above species-in-habitat example, suppose that the marginal qualitative similarity structure on X^2 is described by $T^2 = \{(1, 2, 1)\}^*$, i.e. the smallest TSO containing the triple $(1, 2, 1)$, corresponding to the case in which the habitat "wild" is valued more highly than the habitat "fenced"; thus, the two relevant attributes are "existing at all" ($\{1, 2\}$) and "existing in the wild" ($\{2\}$). Compatibility of $v : 2^{X^1 \times X^2} \rightarrow \mathbf{R}$ with the product is easily seen to be equivalent to the existence of two functions $w_k : 2^{X^1} \rightarrow \mathbf{R}$, $k = 1, 2$, such that for all $S \subseteq X^1 \times X^2$,

$$v(S) = w_1(\text{proj}_1 S) + w_2(\{x^1 \in X^1 : (x^1, 2) \in S\}). \tag{4.1}$$

Intuitively, w_1 gives the value of mere existence of species, while w_2 values existence of species in the wild on top of their mere existence.

4.2 Independent Product

Quantitatively, a natural notion of absence of interaction between dimensions is described as follows.[12] Say that a set function $v : X^1 \times X^2 \rightarrow \mathbf{R}$ is the *independent product* of $v^1 : 2^{X^1} \rightarrow \mathbf{R}$ and $v^2 : 2^{X^2} \rightarrow \mathbf{R}$ if v is separable and for all rectangular sets $S = S^1 \times S^2 \subseteq X^1 \times X^2$,

$$v(S^1 \times S^2) = v^1(S^1) \cdot v^2(S^2). \tag{4.2}$$

That separability and (4.2) indeed uniquely determine a set function v is asserted by the following result.

Theorem 4.1 *Given two set functions* v^1 *and* v^2 *on* X^1 *and* X^2*, respectively, there is a unique separable* $v : 2^{X^1 \times X^2} \rightarrow \mathbf{R}$ *such that* $v(S^1 \times S^2) = v^1(S^1) \cdot v^2(S^2)$ *for all*

[12] All what follows readily extends to the case of the product of an arbitrary (finite) number of coordinates. For expository convenience, we restrict our analysis to the case of two coordinates.

rectangular $S = S^1 \times S^2$. The independent product v is a diversity function if v^1 and v^2 are diversity functions.[13]

Theorem 4.1 is an immediate corollary of the following characterization of the conjugate Moebius inverse (for the definition of the latter, see TD I, Sect. 2.3).

Proposition 4.2 *Let $v : 2^{X^1 \times X^2} \to \mathbf{R}$ be a separable set function. For all rectangular sets $S^1 \times S^2$,*

$$v(S^1 \times S^2) = v^1(S^1) \cdot v^2(S^2)$$

for some set functions $v^k : 2^{X^k} \to \mathbf{R}$ if and only if for all $A^1 \times A^2$,

$$\lambda_{A^1 \times A^2} = \lambda^1_{A^1} \cdot \lambda^2_{A^2},$$

where $\lambda, \lambda^1, \lambda^2$ are the conjugate Moebius inverses of v, v^1, v^2, respectively.

The first part of Theorem 4.1 follows since separability is equivalent to $\lambda_A = 0$ for all non-rectangular sets A. For the second part of Theorem 4.1, recall that diversity functions are characterized by non-negativity of the conjugate Moebius inverse.

To illustrate the notion of independent product in a sociodiversity context (cf. TD I, Sect. 2.5), let X be a set of individuals, and denote by B be the set of books read by someone. For each $b \in B$, let $\emptyset \neq A_b \subseteq X$ denote the readership (extension) of book b. For each $A \subseteq X$, set

$$\lambda_A = \frac{\#\{b : A_b = A\}}{\#B}.$$

Hence, λ_A is the fraction of books read exactly by the set A of individuals; alternatively, λ_A may be interpreted as the probability that a randomly chosen book is read exactly by the individuals in A. For $S \subseteq X$, $v(S) = \lambda(\{A : A \cap S \neq \emptyset\})$ is the fraction of books read by someone in S, or, alternatively, the probability that a randomly chosen book is read by someone in S.

Now suppose that readers $x \in X$ are described by a profile of qualities $x = (x^1, x^2)$, say their age x^1 and their level of education x^2 (measured in years of schooling), so that $X = X^1 \times X^2$. Consider the readership $A_b \subseteq X^1 \times X^2$ of book b. The projection $\text{proj}_1 A_b \subseteq X^1$ can be interpreted as the "readership according to age," in the sense that b appeals to (is read by) someone of age $x^1 \in \text{proj}_1 A_b$; analogously, $\text{proj}_2 A_b \subseteq X^2$ gives the readership according to level of education.

Separability of v, i.e. the condition that $\text{supp}\,\lambda \subseteq \mathcal{A}^{\text{sep}}$, is the requirement that for all $A \in \Lambda$ (equivalently, for all A_b, $b \in B$), $A = \text{proj}_1 A \times \text{proj}_2 A$; hence, a book appeals to a reader if and only if it appeals to her age and her education. In the separable case, a relevant attribute $A \in \Lambda$ can thus be viewed as a multi-dimensional "test": given an individual $x = (x^1, x^2)$, a book passes the test ($x \in A_b$) if and only if it passes the "age"-test ($x^1 \in \text{proj}_1 A_b$) and the "education"-test ($x^2 \in \text{proj}_2 A_b$).

Independence amounts to λ being a product measure, i.e. $\lambda = \lambda^1 \cdot \lambda^2$ for some λ^k on X^k. By Proposition 4.2, this is equivalent to $v(S^1 \times S^2) = v^1(S^1) \cdot v^2(S^2)$ for all S^1, S^2, where $v^k(S^k) := \lambda^k(\{A^k \subseteq X^k : A^k \cap S^k \neq \emptyset\})$. Thus, the probability that someone in $S^1 \times S^2$ reads the randomly chosen book b, i.e. the probability that b passes for some individual in $S^1 \times S^2$ both component tests jointly, equals the probability that

[13] Conversely, if v is a diversity function, either both v^1 and v^2, or both $-v^1$ and $-v^2$, are diversity functions.

b passes the "age"-test for some age $x^1 \in S^1$ times the probability that b passes the "education"-test for some $x^2 \in S^2$. This is intuitive.

Now assume uniformity, i.e. $v(\{x\}) = v(\{y\})$ for all $x, y \in X$, and the normalization $v(\{x\}) = 1$. The following is a general property of the associated quantitative similarity function σ under independence and uniformity.

Fact 4.2 *Suppose that v is an independent product satisfying $v(\{x\}) = 1$ for all $x \in X$. Then,*

$$\sigma((x_1, x_2), (y_1, y_2)) = \sigma((x_1, x_2), (y_1, x_2)) \cdot \sigma((y_1, x_2), (y_1, y_2)).$$

Under the sociodiversity interpretation, quantitative similarity can be interpreted in terms of conditional probability; indeed, $\sigma(x, y)$ is the conditional probability that a randomly chosen book is read by y given that it is read by x. Fact 4.2 asserts that in the independent case these conditional probabilities have a particularly simple multiplicative structure. To illustrate, consider the following three distinguished readers (with at least 15 years of schooling), $x = (50, 15)$, $y = (75, 20)$ and $z = (75, 15)$. The conditional probability that a book read by $(50, 15)$ is also read by $(75, 20)$ equals the conditional probability that a book read by $(50, 15)$ is also read by $(75, 15)$ times the conditional probability that a book read by $(75, 15)$ is also read by $(75, 20)$.

The independent product can be characterized in terms of the underlying preference relation over set-lotteries (cf. TD I, Sect. 2.4). For any fixed $S^1 \in X^1$, a function $v : 2^{X^1 \times X^2} \to \mathbf{R}$ induces a function $v^2_{S^1}$ on 2^{X^2} according to $v^2_{S^1}(\cdot) := v(S^1 \times \cdot)$. Observe that $v^2_{S^1}$ is a diversity function whenever v is a diversity function. Denote by $\succeq^2_{S^1}$ the corresponding von-Neumann-Morgenstern preference on $\Delta^{2^{(X^2)}}$, i.e. the preference over lotteries of subsets of X^2 that arises from taking the expectation of $v^2_{S^1}$. The induced functions $v^1_{S^2}$ and their corresponding preference relations $\succeq^1_{S^2}$ are defined analogously. The following result characterizes functions v on $2^{X^1 \times X^2}$ that satisfy (4.2) as those for which the induced marginal preferences $\succeq^2_{S^1}$ and $\succeq^1_{S^2}$ are independent of S^1 and S^2, respectively.

Theorem 4.2 *Let $v : 2^{X^1 \times X^2} \to \mathbf{R}$ be a set function. There exist two functions $v^1 : 2^{X^1} \to \mathbf{R}$ and $v^2 : 2^{X^2} \to \mathbf{R}$ such that (4.2) holds if and only if, for all non-empty $S^1, W^1 \subseteq X^1$ and all non-empty $S^2, W^2 \subseteq X^2$,*

$$\succeq^2_{S^1} = \succeq^2_{W^1} \quad \textit{and/or} \quad \succeq^1_{S^2} = \succeq^1_{W^2} . \tag{4.3}$$

In the above species-in-habitat example, the diversity function (4.1) is an independent product if and only if the function w_1 is proportional to w_2. In preference terms, this means that, for each set of species $S^1 \subseteq X^1$, the probability π that satisfies

$$\pi \cdot \mathbf{1}_{S^1 \times \{2\}} + (1 - \pi) \cdot \mathbf{1}_\emptyset \sim \mathbf{1}_{S^1 \times \{1\}}$$

does not depend on S^1. I.e., the probability π that makes one indifferent between having all species in S^1 surviving in the wild with probability π (or none at all) and having all species in S^1 surviving in fences and none in the wild is the same for all S^1.

Remark In the literature on non-additive belief representations, the independent product of "belief functions" (conjugate diversity functions) and, more generally, of capacities (see Ghirardato (1997)) has been studied. In particular, Hendon *et al.* (1996)

present a six-fold characterization of the independent product of belief functions, including the product formula for the non-conjugate Moebius inverse which we use in the proof of Proposition 4.2. While the literature assumes throughout (4.2), it does not have a notion of separability to uniquely determine the extension to non-rectangles, nor an analogue to the independence condition (4.3) on induced marginal preferences.

4.3 Application: The Independent Product of Real Lines

We want to apply the analysis to the case of the product of (continuous) real lines. Denote by $\mathcal{F}$ and $\mathcal{K}$ the family of all finite and the family of all compact subsets of $X = \mathbf{R}$, respectively. A monotonic real-valued function v defined on $\mathcal{F}$ can be extended to the domain $\mathcal{K}$ as follows. For all $S \in \mathcal{K}$,

$$v(S) := \sup_{F \in \mathcal{F}, F \subseteq S} v(F),$$

where we allow for the possibility that $v(S) = \infty$. The function v will be called a diversity function if, for all $F \in \mathcal{F}$, the restriction $v|_F$ of v to 2^F is a diversity function on 2^F. Throughout, we assume that v conforms to the line model in the sense that, for any $F \in \mathcal{F}$, $v|_F$ satisfies the Interval Property, i.e. any element of the support of the conjugate Moebius inverse of $v|_F$ is an interval. In this case, $v : \mathcal{K} \to (\mathbf{R} \cup \{\infty\})$ is uniquely determined by $v^{\mathcal{B}}$, the restriction of v to all subsets with at most two elements (cf. TD I, Sect. 4).[14]

As in TD I, Sect. 4.5, the function v will be called translation invariant if, for all S and all $t \in \mathbf{R}$, $v(S) = v(S + t)$, where $S + t := \{x + t : x \in S\}$. In the translation invariant case, v is uniform in the sense that $v(\{x\}) = v(\{y\})$ for all $x, y \in \mathbf{R}$, and we assume the normalization $v(\{x\}) = 1$ throughout. Define a function $f : \mathbf{R} \to \mathbf{R}$ by

$$f(t) := d(0, t) = v(\{0, t\}) - 1.$$

The quantitative similarity relation σ corresponding to v is then given by

$$\sigma(x, y) = 1 - f(|x - y|).$$

The following result is an immediate consequence of TD I, Corollary 4.1.

Fact 4.3 *A translation invariant function v is a diversity function if and only if f is bounded by* 1*, non-decreasing and concave.*

The value of v on compact intervals admits a simple formula as follows.

Fact 4.4 *Let v be a translation invariant diversity function. For all $x \le y$,*

$$v([x, y]) = 1 + f'(0) \cdot (y - x),$$

where $f'(0)$ is the right-hand derivative of f at 0*.*

Observe that, by Fact 4.4, v is finite-valued if and only if $f'(0)$ is finite.[15] Of special interest is the case $f(t) = 1 - e^{-\beta|t|}$ with $\beta > 0$, which we refer to as the *homogeneous*

[14] In fact, it follows from TD I, Th. 4.1, that for each $S \in \mathcal{K}$, $v(S)$ is determined by the values of $v^{\mathcal{B}}$ on all one- and two-element subsets *of* S.

[15] Either of these conditions is equivalent to the ordinal ranking of finite (or compact) sets being continuous in the Vietoris topology as defined in Nehring and Puppe (1996).

translation invariant case (cf. TD I, Sect. 4.5). Let $T_{\mathbf{R}}$ denote the canonical qualitative similarity relation associated to the real line, i.e.

$$(x, y, z) \in T_{\mathbf{R}} \Leftrightarrow [x \geq y \geq z \text{ or } z \geq y \geq x].$$

The following result characterizes the homogeneous translation invariant case in terms of the associated quantitative similarity relation.

Proposition 4.3 *In the translation invariant case, $f(t) = 1 - e^{-\beta|t|}$ if and only if f is non-decreasing and, for all x, y, z,*

$$(x, y, z) \in T_{\mathbf{R}} \Rightarrow \sigma(x, z) = \sigma(x, y) \cdot \sigma(y, z).$$

Now consider the independent product of $\#K$ homogeneous translation invariant lines, i.e. suppose that for any rectangular set $S = \prod_{k \in K} S^k \subseteq \mathbf{R}^K$, one has $v(S) = \prod_{k \in K} v^k(S^k)$, where each v^k is a homogeneous translation invariant diversity function on the real line corresponding to $f^k(t) = 1 - e^{-\beta^k|t|}$.

Fact 4.5 *Let v be the independent product of $\#K$ homogeneous translation invariant diversity functions on the real line, and let σ be the associated quantitative similarity relation. Then,*

$$\sigma(x, y) = e^{-\left(\sum_{k \in K} \beta^k |y^k - x^k|\right)}.$$

Since $d(x, y) = 1 - \sigma(x, y)$, one thus obtains the following picture for the set of all points equi-distant to the origin.

Figure 3: Locus of points equi-distant to 0

In particular, up to multiplicative rescaling of coordinates, there is a unique diversity function on $\mathbf{R}^K$ that is the independent product of homogeneous translation invariant lines.

Let δ denote negative logarithmic quantitative similarity, i.e.

$$\delta(x, y) := -\log \sigma(x, y) = -\log(1 - d(x, y)),$$

and define a corresponding *"logarithmic geodesic betweenness"* T_δ as follows. For all $x, y, z \in \mathbf{R}^K$,

$$(x, y, z) \in T_\delta :\Leftrightarrow \delta(x, y) + \delta(y, z) = \delta(x, z).$$

Proposition 4.4 *Let v be the independent product of homogenous translation invariant real lines. Then, δ is a metric and $T_\delta = \otimes_{k \in K} T_{\mathbf{R}}$.*[16]

[16] The converse statement also holds: If δ is a translation invariant metric and $T_\delta = \otimes_{k \in K} T_{\mathbf{R}}$, then v is the independent product of homogeneous translation invariant real lines.

Similarity is non-Euclidean in two respects here. First, distances are additive only after logarithmic transformation; secondly, the coordinate axes play a distinguished role, in particular, the "circles" of equi-distant points have kinks along the axes. The first difference to the Euclidean paradigm is explained by the general strict subadditivity of dissimilarity, as already discussed in TD I in the context of the line model (cf. TD I, Sect. 4.3). The second is due to the different underlying convex structure. In the independent (separable) product, the convexity is constructed from the component convexities; by contrast, under the Euclidean convexity, all directions are on equal footing, and coordinate axes are chosen as a matter of convention. As a further illustration of the difference in the underlying geometry, consider the value of rectangles in the independent homogeneous product. It follows from Fact 4.4 above that, for all rectangles $\prod_{k\in K}[x^k, y^k]$,

$$v(\prod_{k\in K}[x^k, y^k]) = \prod_{k\in K}(1+\beta^k|y^k - x^k|).$$

Thus, at very small scales (all $|y^k - x^k|$ close to zero), rectangles are ordered approximately according to their "circumference" $\sum_k \beta^k|y^k - x^k|$; at very large scales (all $|y^k - x^k|$ large), on the other hand, rectangles are ordered approximately according to their "volume" $\prod_k |y^k - x^k|$). Thus, contrary to spatial orderings of physical size, diversity comparisons appear to be fundamentally scale-dependent.

5 Conclusion

A key theme of both this paper and TD I and, as we have argued, of any adequate theory of diversity, has been the interrelation of diversity and (dis)similarity. On the multi-attribute approach, the relation is very tight in that diversity can be viewed as the "integral" of point-set dissimilarity, as we argued in TD I, Sect. 2.2. On the other hand, point-set dissimilarity is reducible to point-point dissimilarity using the standard geometric definition of point-set distance only in the case of hierarchical attribute structure (TD I, Sect. 3). In general, for example in the hypercube, point-set dissimilarity is not reducible in this way.

While conventional geometric intuition can mislead at times, geometric concepts and intuitions prove nonetheless to be pervasively helpful. At a qualitative level, we have argued that comparative similarity can be formalized as a betweenness relation that describes the "similarity geometry" of the object space. At a quantitative level, the analogy of the notions of dissimilarity and distance is also helpful, but has to be employed with care. Dissimilarity functions always satisfy the triangle inequality, and are symmetric if and only if all singletons have identical diversity value. While certainly restrictive, the latter assumption is frequently a natural one to make, or may even be entailed by global symmetries (as, e.g., in the case of the translation invariant line). Probably the most significant dis-analogy between dissimilarity and geometric distances is in their link to the underlying qualitative geometry: while geometrically it is almost canonical to require additivity on the line (and, more generally, on "linear betweenness segments"), this condition will only exceptionally be satisfied by dissimilarity functions (TD I, Sect. 4.3). In the context of diversity, the link is weaker: while the dissimilarity metric is always adapted to the underlying qualitative similarity geometry, the qualitative geometry cannot always be recovered in a canonical way from the metric. In special cases, however, we have shown that this can be done: if the underlying convexity is triple-connected, it can be recovered from the metric as "metric betweenness" (Prop. 3.2 above), and in the independent product of homogeneous translation invariant lines as "logarithmic geodesic betweenness" (Prop. 4.4).

Although geometric intuitions are pervasive in the literature on similarity, the extent of their validity within a multi-attribute approach seems quite remarkable in view of the abstract, prima facie non-geometric spirit of that approach.

Appendix: Proofs

Proof of Proposition 2.1 Suppose that $(x, y, z) \notin T_{\mathcal{A}}$, i.e. for some $A' \in \mathcal{A}$, $\{x, z\} \subseteq A'$ and $y \notin A'$. If $\mathcal{A}$ is a hierarchy, there cannot exist a set $A \in \mathcal{A}$ such that $\{x, y\} \subseteq A$ and $z \notin A$. Hence, for all $A \in \mathcal{A}$, $\{x, y\} \subseteq A \Rightarrow z \in A$, i.e. $(x, z, y) \in T_{\mathcal{A}}$. This demonstrates completeness of $T^x_{\mathcal{A}}$, for all x, in the hierarchical case.

Conversely, if $\mathcal{A}$ is not hierarchical, there exist $A, B \in \mathcal{A}$ such that $A \setminus B$, $A \cap B$ and $B \setminus A$ are all non-empty. This implies that for any $y \in A \setminus B$, $x \in A \cap B$, $z \in B \setminus A$ neither $y T^x_{\mathcal{A}} z$, nor $z T^x_{\mathcal{A}} y$.

Proof of Theorem 3.1 in text.

Proof of Theorem 3.2 Let $T \subseteq T_v$; by Theorem 3.1 this is equivalent to $T \subseteq T_\Lambda$. By Fact 2.3, this implies $\mathcal{A}_{T_\Lambda} \subseteq \mathcal{A}_T$. By definition, one has for any family $\mathcal{D} \subseteq 2^X$, $\mathcal{D} \subseteq \mathcal{A}_{T_\mathcal{D}}$. In particular, $\Lambda \subseteq \mathcal{A}_{T_\Lambda}$, hence $\Lambda \subseteq \mathcal{A}_T$.

Conversely, suppose that $\Lambda \subseteq \mathcal{A}_T$. By definition of the associated TSO, this implies $T_{\mathcal{A}_T} \subseteq T_\Lambda$. By Fact 2.3 and Theorem 3.1, $T \subseteq T_\Lambda = T_v$.

Proof of Theorem 3.3 It suffices to show that $T_{\Lambda^*} = T_\Lambda$ since then $\mathcal{A}_{T_{\Lambda^*}} = \mathcal{A}_{T_\Lambda}$ and, by Fact 2.3 and Theorem 3.1, $\mathcal{A}_{T_{\Lambda^*}} = \Lambda^*$ and $T_\Lambda = T_v$, respectively. By definition, $\Lambda \subseteq \Lambda^*$ implies $T_{\Lambda^*} \subseteq T_\Lambda$, hence it remains to show that $T_{\Lambda^*} \supseteq T_\Lambda$. Clearly, $\Lambda \subseteq \mathcal{A}_{T_\Lambda}$, hence $\Lambda^* \subseteq \mathcal{A}_{T_\Lambda}$ since $\mathcal{A}_{T_\Lambda}$ is a CVS. Applying Fact 2.3 twice, one obtains $T_{\Lambda^*} \supseteq T_{\mathcal{A}_{T_\Lambda}} = T_\Lambda$, and hence the desired conclusion.

Proof of Proposition 3.1 Suppose that $\mathcal{A}$ is a hierarchy, and let $\mathcal{A}' \supseteq \{\emptyset, X\}$ be any family such that $T_{\mathcal{A}} = T_{\mathcal{A}'}$. First, observe that any hierarchical family $\mathcal{A}$ containing $\emptyset$ and X is a CVS. Since $\mathcal{A}' \subseteq \mathcal{A}_{T_{\mathcal{A}'}}$, and since $\mathcal{A}$ is a CVS, one obtains by Fact 2.3,

$$\mathcal{A}' \subseteq \mathcal{A}_{T_{\mathcal{A}'}} = \mathcal{A}_{T_{\mathcal{A}}} = \mathcal{A}.$$

Now observe that $\mathcal{A}'$ as a subset of a hierarchy is itself a hierarchy, hence a CVS. Consequently, by a symmetric argument, $\mathcal{A} \subseteq \mathcal{A}'$.

In order to verify the converse implication, let $\mathcal{A}$ be any non-hierarchical family, i.e. there exist $A, B \in \mathcal{A}$ such that $A \setminus B$, $A \cap B$ and $B \setminus A$ are all non-empty. It is easy to verify that in this case the families $\mathcal{A} \cup \{A \cap B\}$ and $\mathcal{A} \setminus \{A \cap B\}$ induce the same TSO according to (2.1), i.e. $T_{\mathcal{A} \cup \{A \cap B\}} = T_{\mathcal{A} \setminus \{A \cap B\}}$.

Proof of Proposition 3.2 It has already been observed in the main text that $T_v \subseteq T_d$. For the converse to hold, triple-connectedness of T_v is clearly necessary. It remains to show that triple-connectedness is also sufficient. Let $(x, y, z) \in T_d$, i.e. $d(x, y) \leq d(x, z)$ and $d(z, y) \leq d(z, x)$, and assume, by way of contradiction, that $(x, y, z) \notin T_v$. By definition of T_v and submodularity of v, one has $d(x, \{y, z\}) < d(x, y)$, and hence

$d(x,\{y,z\}) < d(x,z)$. This implies $(x,z,y) \notin T_v$, and therefore by the symmetry condition T2, $(y,z,x) \notin T_v$. Now $(x,y,z) \notin T_v$ also implies, again by T2, that $(z,y,x) \notin T_v$. A completely symmetric argument as before shows that from this one obtains $(z,x,y) \notin T_v$. Hence, none of (x,y,z), (y,z,x), and (z,x,y) are in T_v, contradicting triple-connectedness of T_v.

Proof of Theorem 3.4 Necessity of the stated conditions is obvious. The proof of sufficiency combines a series of results that have already been established. Right below we show that under the assumptions made, T_d is a TSO. Given this, define $\mathcal{H} := \mathcal{A}_{T_d}$; by Fact 2.3, $T_d = T_{\mathcal{H}}$. Hence, by the assumed completeness of T_d^x and Proposition 2.1, $\mathcal{H}$ is a hierarchy. Since by definition, d is adapted to T_d, and hence also to $T_{\mathcal{H}}$, one can apply the Hierarchy Extension Theorem (TD I, Th. 4.4) to the given $\mathcal{H}$ in order to obtain a unique extension $v : 2^X \to \mathbf{R}$ of $v^{\mathcal{B}}$ such that $\Lambda \subseteq \mathcal{H}$. This demonstrates existence.

To verify uniqueness (of $\mathcal{H}$), let v' be an extension, and let $\mathcal{H}'$ denote the support of its attribute weighting function. By Theorem 3.1, $T_{v'} = T_{\mathcal{H}'}$; by Corollary 3.1, $T_{v'} = T_d$. Hence, $T_{\mathcal{H}'} = T_{\mathcal{H}}$, which by Proposition 3.1 implies $\mathcal{H} = \mathcal{H}'$ up to the inclusion of the universal attribute X. However, the weight of X is uniquely determined by $\lambda_X = \min_{x,y\in X} \sigma(x,y)$, which is non-negative by boundedness of d.

It remains to be shown that T_d is a TSO. Obviously, T_d is reflexive and symmetric, i.e. satisfies T1 and T2, respectively. Hence, we only have to establish the transitivity condition T3. This is done in two steps. First, we show that completeness of T_d^x implies (standard) transitivity of T_d^x, i.e. all relations T_d^x are weak orders. We then show that, for any symmetric ternary relation T such that all T^x are weak orders, T satisfies the transitivity condition T3.

Hence, suppose that $(x,y,z) \in T_d$ and $(x,z,w) \in T_d$. We have to show that $(x,y,w) \in T_d$. Using (3.4), one has $\sigma(x,w) \leq \sigma(x,z)$ (from $(x,z,w) \in T_d$) and $\sigma(x,z) \leq \sigma(x,y)$ (from $(x,y,z) \in T_d$), hence $\sigma(x,w) \leq \sigma(x,y)$. Now assume that $(x,y,w) \notin T_d$; by completeness of T^x, this implies $(x,w,y) \in T_d$, in particular $\sigma(x,y) \leq \sigma(w,y) = \sigma(y,w)$. Thus, $\sigma(x,w) \leq \sigma(x,y)$ and $\sigma(x,w) \leq \sigma(y,w)$, which by (3.4) implies $(x,y,w) \in T_d$, a contradiction.

We now show that symmetry, completeness and transitivity of all T^x together imply transitivity of T. Take x,x',y,z,z' such that $(x,x',z) \in T$, $(x,z',z) \in T$ and $(x',y,z') \in T$. By completeness, $(x,x',z') \in T$ or $(x,z',x') \in T$; without loss of generality, assume $(x,x',z') \in T$. By symmetry, $(z',x',x) \in T$ as well as $(z',y,x') \in T$. By transitivity of $T^{z'}$, $(z',y,x) \in T$, hence by symmetry, $(x,y,z') \in T$. Finally, by transitivity of T^x, $(x,y,z) \in T$.

Proof of Theorem 3.5 We only prove sufficiency of the stated conditions. By Krantz, Luce, Suppes and Tversky (1979), a ternary relation T can be represented as $T = T_{\mathcal{L}}$ for $\mathcal{L}$ associated to some unique (up to reversal) linear order $\geq$ on X if and only if T satisfies the following five conditions. (i) symmetry, (ii) triple-connectedness, (iii) antisymmetry (in the sense that $[(x,y,z) \in T$ and $(x,z,y) \in T] \Rightarrow y = z)$, (iv) line-transitivity, and (v) (standard) transitivity of T^x for all x. We show that under the assumptions stated in Theorem 3.5, T_d satisfies all five conditions. It has already been observed in the main text that T_d satisfies (i) and (ii). Condition (iv) holds by assumption. To verify (iii), assume by way of contradiction, that $(x,y,z) \in T_d$, $(x,z,y) \in T_d$ and $y \neq z$. By symmetry, $(y,z,x) \in T_d$, hence by line-transitivity, $(x,y,x) \in T_d$. Since d is adapted to T_d, $d(x,y) \leq d(x,x) = 0$. By strict positivity of d, this implies $y = x$

and hence, by $(x,z,x) \in T_d$, also $z = x$, the desired contradiction.

In order to verify (v), observe first that for antisymmetric T_d, (3.4) can be strengthened to

$$(x,y,z) \in T_d \Leftrightarrow \sigma(x,z) < \min\{\sigma(x,y), \sigma(y,z)\},$$

whenever $y \neq x,z$. Now let $(x,y,z) \in T_d$ and $(x,z,w) \in T_d$, and assume without loss of generality that x,y,z,w are pairwise different. We derive a contradiction from the assumption that $(x,y,w) \notin T_d$. By triple-connectedness, either $(y,w,x) \in T_d$ or $(w,x,y) \in T_d$. In the former case, one obtains using symmetry, $\sigma(x,y) < \sigma(x,w) < \sigma(x,z) < \sigma(x,y)$, a contradiction. In the latter case, one obtains by line-transitivity (applied to (w,x,y) and (x,y,z)) that $(w,x,z) \in T_d$. Hence, $\sigma(w,z) < \sigma(w,x) = \sigma(x,w) < \sigma(z,w) = \sigma(w,z)$, again a contradiction. This shows (v).

Let $\geq$ be the linear order on X such that $T_d = T_{\mathcal{L}}$. The dissimilarity metric d satisfies all requirements in order to apply the Line Extension Theorem 4.3 of TD I. Hence, there exists a unique extension of $v^{\mathcal{B}}$ to a diversity function v on X that satisfies the Interval Property with respect to $\geq$. By Corollary 3.1, $T_v = T_d = T_{\mathcal{L}}$.

Proof of Proposition 4.1 It is clear that the relation $\hat{T}$ defined by

$$(x,y,z) \in \hat{T} :\Leftrightarrow [\text{ for all } k : (x^k, y^k, z^k) \in T^k],$$

is a TSO that contains both, T^{weak} and T^{sep}.

Conversely, suppose that $(x,y,z) \in \hat{T}$, and denote by $\tilde{T}(x,z)$ the segment spanned by x and z with respect to the product similarity, $\tilde{T}(x,z) := \{x' : (x,x',z) \in \otimes_k T_k\}$. We show that $y \in \tilde{T}(x,z)$. For simplicity, we assume two coordinates only; the general case follows along the same lines. Since the product similarity contains T^{sep}, one obtains $(x^1,z^2) \in \tilde{T}(x,z)$ and $(z^1,x^2) \in \tilde{T}(x,z)$. Since the product similarity contains T^{weak}, one has $((x^1,w),(y^1,w),(z^1,w)) \in \otimes_k T^k$ for $w = x^2$ and $w = z^2$, and $((y^1,x^2),(y^1,y^2),(y^1,z^2)) \in \otimes_k T^k$. Using this, repeated application of T3 yields $y \in \tilde{T}(x,z)$, i.e. $(x,y,z) \in \otimes_k T^k$.

Proof of Fact 4.1 The proof uses the following result which is well-known in the literature on abstract convexity theory (see Jamison (1974), van de Vel (1993, p. 87)). For any family $\{\mathcal{A}^k\}_{k\in K}$ of CVSs, the product

$$\otimes_{k\in K}\mathcal{A}^k := \{A : A = \prod_{k\in K} A^k, A^k \in \mathcal{A}^k\}$$

is a CVS.

By Proposition 4.1, $\otimes_k T^k = T_{(\otimes_k \mathcal{A}^k)}$, where $\mathcal{A}^k = \mathcal{A}_{T^k}$ is the CVS corresponding to T^k. Hence, $\mathcal{A}_{(\otimes_k T^k)} = \mathcal{A}_{T_{(\otimes_k \mathcal{A}^k)}}$. By the above result and Fact 2.3, $\mathcal{A}_{T_{(\otimes_k \mathcal{A}^k)}} = \otimes_k \mathcal{A}^k$.

Proof of Proposition 4.2 Let v be separable and suppose that $v(S^1 \times S^2) = v^1(S^1)\cdot v^2(S^2)$ for some functions v^1, v^2 and all S^1, S^2. We use the following formula which is based on a standard result in combinatorics (see, e.g. Cameron (1994, Prop. 12.7.5, p.201), Hendon *et al.* (1996)). For any rectangular $A = \prod_k A^k$, the value of the conjugate Moebius inverse λ of a separable v at A is given by

$$\lambda_A = \sum_{B=\prod_k B^k \subseteq A} (-1)^{\sum_k (\#A^k - \#B^k)} \cdot \hat{v}(B), \qquad \text{(A.1)}$$

where $\hat{v}(B) := v(X) - v(B^c)$ denotes the *loss function* associated to v (recall that B^c denotes the complement of B in X). Furthermore, for all $B^1 \times B^2 \subseteq X^1 \times X^2$,

$$v((B^1 \times B^2)^c) = v(X^1 \times (B^2)^c) + v((B^1)^c \times X^2) - v((B^1)^c \times (B^2)^c). \tag{A.2}$$

Indeed, (A.2) is easily verified by considering the conjugate Moebius inverse λ of v and distinguishing $A^1 \times A^2 \in \Lambda$ according to whether for $k = 1, 2$, $A^k \cap (B^k)^c$ is empty or not. Using (A.1) and (A.2), one thus obtains for all $A^1 \times A^2$,

$$\begin{aligned}
&\lambda_{A^1\times A^2} \\
&= \sum_{B^1\times B^2 \subseteq A^1\times A^2} (-1)^{\sum_k (\#A^k - \#B^k)} \hat{v}(B^1 \times B^2) \\
&= \sum_{B^1\times B^2 \subseteq A^1\times A^2} (-1)^{\sum_k (\#A^k - \#B^k)} \left[v(X^1 \times X^2) - v((B^1 \times B^2)^c)\right] \\
&= \sum_{B^1\times B^2 \subseteq A^1\times A^2} (-1)^{(\#A^1 - \#B^1)} \cdot (-1)^{(\#A^2 - \#B^2)} \left[v^1(X^1) \cdot v^2(X^2)\right. \\
&\qquad \left. - v^1(X^1) \cdot v^2((B^2)^c) - v^1((B^1)^c) \cdot v^2(X^2) + v^1((B^1)^c) \cdot v^2((B^2)^c)\right] \\
&= \sum_{B^1 \subseteq A^1, B^2 \subseteq A^2} (-1)^{(\#A^1 - \#B^1)} \cdot (-1)^{(\#A^2 - \#B^2)} \left[v^1(X^1) - v^1((B^1)^c)\right] \\
&\qquad \cdot \left[v^2(X^2) - v^2((B^2)^c)\right] \\
&= \left(\sum_{B^1 \subseteq A^1} (-1)^{(\#A^1 - \#B^1)} \hat{v}^1(B^1)\right) \cdot \left(\sum_{B^2 \subseteq A^2} (-1)^{(\#A^2 - \#B^2)} \hat{v}^2(B^2)\right) \\
&= \lambda^1_{A^1} \cdot \lambda^2_{A^2}.
\end{aligned}$$

Conversely, it is straightforward to verify that any separable diversity function $v : 2^{X^1 \times X^2} \to \mathbf{R}$ defined from v^1 and v^2 according to $\lambda_{A^1\times A^2} := \lambda^1_{A^1} \cdot \lambda^2_{A^2}$ has the required product form.

Proof of Fact 4.2 Denote by $\sigma^k(x^k, y^k) = \lambda^k(\{A^k : \{x^k, y^k\} \subseteq A^k\}$. By Proposition 4.2, $\sigma(x, y) = \sigma^1(x^1, y^1) \cdot \sigma^2(x^2, y^2)$. From this, the claim follows at once since $v(\{x\}) = \sigma(x, x) = 1$.

Proof of Theorem 4.2 Suppose that $v : 2^{X^1 \times X^2} \to \mathbf{R}$ is normalized so that $v(\emptyset) = 0$. If v satisfies (4.2), $v^2_{S^1}(\cdot) := v(S^1 \times \cdot) = v^1(S^1) \cdot v^2(\cdot)$. Hence, for any non-empty S^1, W^1, $v^2_{S^1}$ equals $v^2_{W^1}$ up to multiplication by a positive scalar. Consequently, $v^2_{S^1}$ and $v^2_{W^1}$ represent the same von-Neumann-Morgenstern preference on $\Delta^{2^{(X^2)}}$. An analogous argument shows that $\succeq^1_{S^2} = \succeq^1_{W^2}$.

Conversely, suppose that for all non-empty S^1, W^1, the preferences $\succeq^2_{S^1}$ and $\succeq^2_{W^1}$ coincide. This implies that for all S^1, W^1, $v^2_{S^1}$ equals $v^2_{W^1}$ up to multiplication by a positive scalar. Define a function $v^1 : 2^{X^1} \to \mathbf{R}$ as follows. For $S^1 \in X^1$,

$$v^1(S^1) := \frac{v^2_{S^1}(S^2)}{v^2_{X^1}(S^2)}. \tag{A.3}$$

Observe that the thus defined v^1 does not depend on the choice of S^2 in (A.3). By construction, $v(S^1 \times S^2) = v^1(S^1) \cdot v^2_{X^1}(S^2)$, for all S^1, S^2.

Proof of Fact 4.4 Without loss of generality, take $x = 0$. Consider, for each n, the set $S_n := \{l \cdot \frac{y}{n} : l = 0, 1, ..., n\}$. From the Line equation (cf. TD I, (4.1)) it follows that $v(S_n) = 1 + n \cdot f(y/n)$. Hence,

$$v([0, y]) \geq \lim_{n \to \infty} 1 + n \cdot f(y/n) = 1 + f'(0) \cdot y.$$

Observe that the limit exists because f is concave (it may take the value ∞). On the other hand, $v([0, y]) \leq \lim_n 1 + n \cdot f(y/n)$, because for any finite subset $F = \{x_1, ..., x_m\} \subseteq [0, y]$ with $x_1 < x_2 < ... < x_m$,

$$v(F) = 1 + \sum_{i=2}^{m} f(x_i - x_{i-1}) \leq 1 + f'(0) \sum_{i=2}^{m} (x_i - x_{i-1}) = 1 + f'(0) \cdot (x_m - x_1),$$

by concavity of f.

Proof of Proposition 4.3 Define a function $g : \mathbf{R}_+ \to \mathbf{R}$ by $g(t) := \log \sigma(x, y)$, where $|y - x| = t$. The stated condition on σ implies $g(t + t') = g(t) + g(t')$ for all t, t'. By Aczél (1966, Th. 1, p.34), g must be linear.

Proof of Fact 4.5 As in the proof of Fact 4.2, Proposition 4.2 implies

$$\sigma(x, y) = \lambda(\{\prod_k A^k : \{x, y\} \subseteq \prod_k A^k\}) = \prod_k \lambda^k(\{A^k : \{x^k, y^k\} \subseteq A^k\}) = \prod_k \sigma^k(x^k, y^k),$$

which immediately implies the desired result.

Proof of Proposition 4.4 The proof is straightforward, noting that in the homogeneous case, $\delta(x, y) = \sum_k \beta^k |y^k - x^k|$ by Fact 4.5.

References

[1] Aczél, J. (1966), *Lectures on Functional Equations and their Applications*, Academic Press, New York.

[2] Benzecri, J.P. *et.al.* (1973), *L'analyse des Données*, vol. 1: La Taxinomie, Dunod, Paris.

[3] Cameron, P.J. (1994), *Combinatorics*, Cambridge University Press, Cambridge.

[4] Fishburn, P.C. (1985), *Interval Orders*, John Wiley, New York.

[5] Gärdenfors, P. (1990), Induction, Conceptual Spaces and AI, *Philosophy of Science* **57**, 78-95.

[6] Ghirardato, P. (1997), On Independence for Non-Additive Measures, with a Fubini Theorem, *Journal of Economic Theory* **73**, 261-292.

[7] Hendon, E., H.J. Jacobsen, B. Sloth and T. Tranaes (1996), The Product of Capacities and Belief Functions, *Mathematical Social Sciences* **32**, 95-108.

[8] Hilbert, D. (1899), *Grundlagen der Geometrie*, Teubner, Leipzig.

[9] Jamison, R.E. (1974), A General Theory of Convexity, Dissertation, University of Washington, Seattle, Washington.

[10] Johnson, S.C. (1967), Hierarchical Clustering Schemes, *Psychometrika* **32**, 241-254.

[11] Krantz, D.H., R.D. Luce, P. Suppes and A. Tversky (1979), *Foundations of Measurement*, vol. 2, Academic Press, New York.

[12] Menger, K (1928), Untersuchungen über die Allgemeine Metrik, *Mathematische Annalen* **100**, 75-163.

[13] Nehring, K. (1997), A Theory of Qualitative Similarity, *mimeo.*

[14] Nehring, K. and C. Puppe (1996), Continuous Extensions of an Order on a Set to the Power Set, *Journal of Economic Theory* **68**, 456-479.

[15] Nehring, K. and C. Puppe (1999a), A Theory of Diversity I: Consulting Noah, *mimeo.*

[16] Nehring, K. and C. Puppe (1999b), Modelling Economies of Scope, *mimeo.*

[17] Pasch, M. (1882), *Vorlesungen über die neuere Geometrie*, Teubner, Leipzig.

[18] Shafer, G. (1976), *A Mathematical Theory of Evidence*, Princeton University Press, Princeton.

[19] Suppes, P. (1972), Some Open Problems in the Philosophy of Space and Time, *Synthese* **24**, 298-316.

[20] van de Vel, M. (1993), *Theory of Convex Structures*, North-Holland, Amsterdam.