Diversity in a fuzzy world: a review of models and measures

Carlo Ricotta[1], Andrea Campagner[2], Sandrine Pavoine[3], Davide Ciucci[2]

[1]Department of Environmental Biology, University of Rome 'La Sapienza', Rome, Italy; [2]Department of Informatics, Systems and Communication, University of Milano-Bicocca, Milano, Italy; [3]Centre d'Ecologie et des Sciences de la Conservation (CESCO), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, Paris, France.

✉ Carlo Ricotta (carlo.ricotta@uniroma1.it)

ORCID
Carlo Ricotta https://orcid.org/0000-0003-0818-3959
Andrea Campagner https://orcid.org/0000-0002-0027-5157
Sandrine Pavoine https://orcid.org/0000-0003-2767-6484
Davide Ciucci https://orcid.org/0000-0002-8083-7809

Abstract

Biological diversity metrics have evolved from simple species counts to measures that incorporate increasingly complex functional information. A critical aspect is that different families of diversity indices are grounded in distinct mathematical models for representing uncertainty, including classical set theory, probability theory, and fuzzy set theory, which shape what these measures can reveal about community structure. In this paper, we propose a unified theoretical framework that links commonly used diversity indices to their underlying models of uncertainty, thereby clarifying the ecological and mathematical meaning of the information they summarize. As one of our main results, we show that Rao's quadratic diversity, one of the best known indices of functional diversity, can be interpreted as a measure of strife or discord for a set of functionally overlapping species. We further demonstrate that Rao's diversity can be expressed as the mathematical expectation of the fuzzy specificity (or functional distinctiveness) of the individual species in the community, thereby revealing a closer ecological and statistical connection between these two forms of uncertainty than previously recognized. Overall, our approach provides an integrated perspective on the relationship between uncertainty theory and biodiversity measurement, offering new insights for understanding ecological processes and for developing more coherent and informative biodiversity indices.

Keywords: Functional distinctiveness; Fuzzy sets; Information theory; Rao quadratic diversity; Shannon entropy; Simpson diversity; Species richness; Specificity; Uncertainty

Introduction

In community ecology, diversity measures are fundamental tools for understanding the mechanisms shaping community structure and composition over varying spatial scales. Fisher et al. (1943) were the first to use the term 'index of diversity' to refer to the shape parameter of a log-series distribution describing how individuals are distributed among species within a community (what is now known as Fisher's alpha). A few years later, Simpson (1949) proposed replacing diversity indices based on characteristic parameters of species-abundance distributions with the type of diversity indices we use today, namely 'general parameters' in the sense of Good (1953) that measure the heterogeneity of a species assemblage independently of any specific form of abundance distribution. For roughly the next fifty years, these distribution-free diversity measures were calculated exclusively from species relative abundances without considering other attributes of community structure, such as functional differences between species.

In the specific case of the Simpson index, given a community C composed of N species, let

K_i $(i = 1, 2, ..., N)$ be the number of individuals belonging to species *i* such that $\sum_{i=1}^{N} K_i = K$, and let $p_i = K_i / K$ denote the relative abundance of *i*. The Simpson index of diversity is then defined as the probability that two individuals randomly selected from C do not belong to the same species:

$$S = 1 - \sum_{i=1}^{N} p_i^2 \qquad (1)$$

Traditional diversity measures, like the Simpson index, typically summarize community diversity based solely on species commonness and rarity, implicitly assuming that all species are equally and maximally distinct from each other (Mouchet et al. 2010; Ricotta et al. 2023). However, in a more contemporary view of community ecology, it is widely accepted that the species ecological strategies are linked to their functional traits. Therefore, from a functional perspective, the species may overlap to varying degrees, depending on differences in their functional traits. This led Díaz and Cabido (2001) to state that "functional diversity, or the value and range of species traits, rather than species numbers per se, strongly determines ecosystem functioning."

A few years earlier, in an equally important but much less celebrated paper, Izsák and Papp (1995) were the first to propose measuring functional diversity using an index now generally known as Rao's (1982) quadratic diversity. While the conventional abundance-based diversity measures commonly used at that time ignored functional or other ecologically relevant differences among species, Rao (1982) was among the first to introduce a diversity index that explicitly incorporates such differences (see also Hendrickson and Ehrlich 1971). By integrating species abundances with their functional traits, Rao's quadratic diversity offers a more comprehensive understanding of the relationships between plant diversity and ecosystem processes (Pavoine 2012; Ricotta et al. 2021).

This fundamental shift in perspective has led to the development of numerous functional diversity indices that integrate in different ways species abundance with their functional traits (Walker et al. 1999; Villéger et al. 2008; Schmera et al. 2009; Laliberté and Legendre 2010; Chao et al. 2014; Ricotta et al. 2016). As a result, there are now literally thousands of studies linking functional diversity to various dimensions of community functioning.

A critical aspect is that different families of diversity indices, from simple species counts to more recent measures of functional diversity, are grounded in distinct mathematical frameworks for representing the uncertainty associated with specific aspects of community structure, including classical set theory, probability theory, and fuzzy set theory (Ricotta and Pavoine 2025). These models implicitly shape how species and communities are represented. Consequently, the choice of a particular model ultimately constrains what diversity indices can or cannot reveal about community structure.

The aim of this paper is thus to relate the diversity indices commonly used in community ecology to their underlying mathematical models of uncertainty. The main goal is to provide a comprehensive theoretical framework that explains what each measure captures and how they interrelate, thereby ensuring that they really measure what we want to measure (Sherwin et al. 2006).

Methods

Diversity and uncertainty

Biological communities are highly complex multi-dimensional systems. Therefore, their investigation requires multivariate methods of exploratory data analysis. The primary purpose of these analyses is to describe biological patterns and their interrelationships, whereas estimation and statistical inference generally assume a secondary role (Podani 2000). From the standpoint of a community ecologist, diversity measures may be thus conceived as a class of multivariate summary statistics designed to capture distinct dimensions of compositional uncertainty, including species richness, abundance distributions, and functional traits (Solow and Polasky 1994; Ricotta 2005; Ricotta and Feoli 2024). In this view, although many of the most widely used diversity measures originated in very different contexts and were developed for distinct purposes, they have all been used to summarize the uncertainty associated with various facets of community structure, without any inferential power or direct mechanistic or biological meaning: high diversity simply implies high unpredictability, and vice versa.

Therefore, contrary to those who argue that diversity as an ecological concept is fundamentally distinct from diversity as a statistical index (see Ricotta 2005), we concur with Molinari (1989) that: "Diversity as an unequivocal and inherent property of a biological system does not exist. Instead, as it is the case with any other statistic, diversity values are merely numbers and their relevance to ecological problems must be judged [*a posteriori*] on the basis of observed correlations with other environmental variables." Clearly, the nature of diversity depends on the mathematical framework within which the index is formalized. Each formalization relies on a model designed to capture the uncertainty associated with specific aspects of community structure. Consequently, adopting a specific mathematical model necessarily constrains the analysis to the assumptions and limitations inherent in that model (Klir and Wierman 1999).

The following sections review the principal uncertainty-based diversity indices commonly used in ecological studies, focusing on their underlying mathematical models arranged in order of increasing complexity.

Species richness

Species richness is the most fundamental and commonly used measure of biological diversity. It refers to the number of distinct species present in a community (or assemblage, collection, etc.), regardless of their abundance. In terms of uncertainty formalizations, species richness refers to classical set theory, which is probably the simplest way to express uncertainty. In its most basic representation, a classical, or crisp set C (i.e., the community) is a finite collection of *N* discrete elements (species). The boundaries of a crisp set are sharply defined: an element either fully belongs to the set, or it does not (Figure 1). Intuitively, this binary membership means that the uncertainty associated with the set depends only on the number of its elements. Therefore, a community with a greater number of species is considered more diverse (i.e., more unpredictable) than one with fewer species.

The measurement of uncertainty in finite crisp sets was first introduced by Hartley in (1928). This type of uncertainty arises when we know that a certain alternative of interest belongs to a particular set of alternatives, but we do not know which one it is. To identify the specific alternative, information is required to eliminate this uncertainty. Therefore, the amount of uncertainty associated with a set of alternatives can be measured by the amount of information needed to remove it. More generally, the rationale underlying the relationship between uncertainty and information is that the amount of information gained from observing the outcome of an experiment depending on chance can be regarded as equivalent to the amount of uncertainty associated with this outcome before the experiment was carried out (Rényi 1970).

According to Hartley (1928), let us consider a finite set containing *N* elements. Sequences can be formed from the *N* elements by successive selections. If *s* selections are performed, then N^s different sequences can potentially be obtained. Assuming that the amount of information $I(N^s)$ required to remove the uncertainty associated with the *s* selections is proportional to *s*, we obtain:

$$I(N^s) = \kappa(N) \times s \tag{2}$$

where $\kappa(N)$ is a constant that depends on *N*.

Given two sets with *N* and *M* elements, respectively, if we perform s_1 selections from the first set and s_2 selections from the second set, and the number of resultant sequences is the same in both cases, then the amount of information should also be the same. Formally, if

$$N^{s_1} = M^{s_2}$$

then

$$\kappa(N)\times s_1=\kappa(M)\times s_2 \tag{3}$$

from which

$$\frac{\kappa(N)}{\log_b N}=\frac{\kappa(M)}{\log_b M} \tag{4}$$

This equation can only be satisfied if

$$\kappa(N)=c\log_b N$$

where b and c are positive constants ($b>1,\ c>0$) that determine the unit in which uncertainty is measured. For $b=2$ and $c=1$, which is the standard choice in information theory, uncertainty is measured in bits.

Substituting the value of κ in Eq. (2), we obtain:

$$I(N^s)=s\log_2 N=\log_2 N^s \tag{5}$$

That is, for a finite crisp set, the information associated with s sequences of its N elements is equal to the logarithm of the number of possible sequences. Therefore, setting $s=1$, the information associated with a single selection of equally likely outcomes is the logarithm of the number of available elements N:

$$I(N)=\log_2 N \tag{6}$$

Eq. (6) is commonly referred to as Hartley's formula. Note that, aside from information theory, where information content is typically expressed in bits, ecological applications generally use the natural logarithm or the base-10 logarithm.

According to Klir and Wierman (1999), uncertainty expressed in terms of sets of alternatives arises from the nonspecificity present in each set. The larger the set, the less specific the resulting prediction, and full specificity is achieved when only one alternative is possible. Therefore, the uncertainty conveyed by sets of possible alternatives is well described by the term *nonspecificity*.

Rényi (1970) showed that Hartley's formula satisfies a set of properties, the most important of which is additivity, thereby justifying its interpretation as a measure of information associated with selecting one element at random from a set of N equally probable choices (for details, see Rényi 1970). By relaxing the requirement of additivity, a condition that has been invoked only rarely in community ecology (e.g., Juhász-Nagy and Podani 1983; Juhász-Nagy 1993), when information about the number of individuals per species is unavailable, or when species abundances are deliberately ignored, it becomes natural to measure the diversity of a biological community simply using its number of species N.

Abundance-based diversity: The Shannon and the Simpson formula

Apart from the Simpson diversity, the other fundamental measure of diversity that is based on species abundances is the Shannon entropy (H), which was first introduced in an ecological context by Good (1953) to summarize the diversity "of a population of animals of various species" and later popularized in numerous studies, particularly by MacArthur (1955, 1961) and Margalef (1958).

Let C be a universal set (a sample space, or, in ecological terms, a community) containing K elements (individuals) belonging to N crisp subsets (the species). Each individual either fully belongs to a given species, or it does not (Figure 1). Therefore, the species are considered mutually exclusive (one individual cannot simultaneously belong to two or more species) and exhaustive (all individuals belong to one of the N species in C), i.e., they form a partition.

According to Rényi (1970), the necessary amount of information required to fully characterize a single individual in C is $\log_2 K$ (i.e., Hartley's formula applied to the K individuals in C). However, if we already know that an individual in C belongs to species i, we still need $\log_2 K_i$ units of information to fully characterize it, where K_i is the number of individuals of species i.

Therefore, on average, the information required to characterize an individual in C, once its species is already known, is:

$$\sum_{i=1}^{N} \frac{K_i}{K} \log_2 K_i = \sum_{i=1}^{N} p_i \log_2 K p_i \tag{7}$$

where $p_i = K_i / K$ is the probability of randomly drawing an element of species i from the set C, with $\sum_{i=1}^{N} p_i = 1$.

It follows that the total information required to characterize one individual in C consists of two parts: the first part, H, determines the species to which the individual belongs, and the second part, $\sum_{i=1}^{N} p_i \log_2 N p_i$, characterizes the individual among those belonging to that species:

$$\log_2 K = H + \sum_{i=1}^{N} p_i \log_2 K p_i \tag{8}$$

which leads to Shannon's (1948) formula:

$$H = \sum_{i=1}^{N} p_i \log_2 \frac{1}{p_i} \tag{9}$$

Eq. 9 is commonly referred to as Shannon's entropy and expresses the probabilistic uncertainty in assigning a randomly selected individual to one of the N species in C, assuming that the species are exhaustive and mutually exclusive.

As with Simpson's index, the value of the Shannon entropy increases with the number of species in C and with the evenness of their relative abundances. In both cases, for a non-empty community, diversity is zero if the community contains only one species and progressively increases with increasing species richness and evenness. For a community composed of N species, diversity reaches its maximum when all species are equally abundant. (i.e., if $p_i = 1/N$ for all $i = 1, 2, ..., N$). Specifically, for an equiprobable distribution of species abundances, the Shannon index reduces to the Hartley formula $H = \log_2 N$, whereas the Simpson index simplifies to $S = 1 - 1/N$ (i.e., an increasing function of species richness). This is intuitively reasonable: in a probabilistic framework, uncertainty is maximal when all events have the same probability (i.e., when the probability of drawing an individual of a given species is the same for all species). Therefore, contrary to those who consider evenness a problematic component of diversity theory (see e.g., Alroy 2025), the increase of diversity with increasing evenness is a *mathematical necessity* (*sensu* Hurlbert 1971), arising from the relationship between diversity and probabilistic uncertainty. Given their probabilistic nature, both indices can serve as community-level indicators of probabilistic uncertainty (also referred to as *strife* or *discord*), reflecting the disagreement in selecting among several mutually exclusive alternatives (Klir and Wierman 1999). High values of Shannon entropy or Simpson diversity indicate a high level of uncertainty regarding the specific identity of a randomly sampled individual.

Note that, since the Simpson index and the Shannon entropy are both expressed in terms of species relative abundances p_i, their application is not restricted to discrete counts of individuals. They can also be extended to continuous measures of quantity, such as species cover or biomass, which are equally important ecological variables, particularly in vegetation science where counts of individuals are often meaningless or impossible, as in the case of clonal plants (Hill 1973).

In terms of diversity formalization, both measures fall within the definition proposed by Patil and Taillie (1982), who defined diversity as the expected (i.e., average) rarity of the relative abundances of the species in C:

$$D = \sum_{i=1}^{N} p_i \times \rho(p_i) \tag{10}$$

In their formulation, Patil and Taillie (1982) expressed the rarity $\rho(p_i)$ of species *i* as a non-negative, monotonically decreasing function that depends solely on its relative abundance p_i. The function is defined over the interval $(0,1]$, such that when $p_i = 1$, $\rho(p_i) = 0$, implying that a community consisting of a single species has zero diversity, whereas for $p_i = 0$, the function is undefined, implying that species not present in the community do not contribute to its diversity. Consequently, the resulting measure of diversity is determined by the specific form of the rarity function.

In terms of average rarity, the Simpson index can be expressed as:

$$S = \sum_{i=1}^{N} p_i (1 - p_i) \tag{11}$$

where the rarity function $\rho_S(p_i) = (1 - p_i)$ decreases linearly with species abundance. Similarly, the rarity function underlying Shannon entropy can be expressed as: $\rho_H(p_i) = \log_2(1/p_i) = -\log_2 p_i$. Hence, both indices can be placed under the same formal umbrella. The primary difference lies in their sensitivity to species abundances: Shannon entropy is more responsive to changes in the abundance of rare species, whereas the Simpson index is more sensitive to variations in the abundance of common species and is largely unaffected by changes in the abundance of rare ones (Hill 1973; Peet 1974).

In the context of strife or discord, recall that when species are considered maximally distinct, p_i denotes the probability that an individual randomly drawn from C belongs to species *i*. Consequently, due to its linearity, Simpson's rarity function $(1 - p_i) = \sum_{j \neq i}^{N} p_j$ summarizes the probability that an individual randomly selected from C does not belong to species *i*. That is, the complement of p_i. Therefore, in terms of species abundances, ρ_S can be also interpreted as the total abundance of all species that conflict with species *i*. Likewise, the rarity function associated with the Shannon entropy $-\log\left(1 - \sum_{j \neq i}^{N} p_j\right)$ summarizes the conflict of all species that differ from *i* in the range $[0, \infty]$. Accordingly, conventional abundance-based diversity measures such as Shannon and Simpson diversity can both be viewed as measures of expected conflict among distinct species within a community (Klir and Wierman 1999; Ricotta and Szeidl 2006).

Trait-based diversity: Rao's quadratic diversity

Whereas traditional diversity metrics primarily account for species richness and abundance, it is broadly acknowledged that a deeper understanding of ecosystem functioning arises from examining the range and distribution of the species' functional traits. Typically, the data used to describe the functional structure of a community include a set of functional traits associated with each species. Since many diversity indices depend on measures of functional dissimilarity, these traits are commonly transformed into a square matrix representing multivariate functional dissimilarities between all pairs of species (Villéger et al. 2008).

Using data on species abundances and their functional dissimilarities, several studies have demonstrated that Rao's (1982) quadratic diversity provides an effective way to quantify the functional diversity of a community. Quadratic diversity represents the expected functional dissimilarity between two individuals randomly selected (with replacement) from the community:

$$Q = \sum_{i,j}^{N} p_i p_j d_{ij} \tag{12}$$

where d_{ij} denotes the symmetric dissimilarity between species *i* and *j* with $d_{ij} = d_{ji}$ and $d_{ii} = 0$.

Although the use of functional diversity indices largely depends on current research trends and the specific questions being addressed, among the various existing measures of functional diversity, Rao's quadratic diversity remains by far the index with the most robust and well-studied theoretical foundation (see e.g., Shimatani 2001; Champely and Chessel 2002; Rao 2010; Pavoine 2012; Ricotta et al. 2016, 2022). According to Eq. (12), the value of Q increases with increasing functional dissimilarity among species. Therefore, Ricotta et al. (2021) proposed this index as an indicator of community-level functional uniqueness, reflecting the extent to which individuals of a given species exhibit distinctive functional traits not shared with other species in the community. Note that, as with conventional abundance-based measures, the calculation of Rao's Q is not limited to discrete counts of individuals but can also be applied to continuous ecological variables.

For the purposes of this study, we assume that the pairwise dissimilarities between species d_{ij} are bounded within the range $[0,1]$. Consequently, their complement $s_{ij} = 1 - d_{ij}$ can be considered a measure of functional similarity between species *i* and *j*.

In set-theoretical terms, the idea that species are not maximally dissimilar but share a certain degree of overlap determined by their functional traits conforms to the assumptions of fuzzy set theory. Unlike a crisp set, in which an element either fully belongs to the set or does not, the elements of a fuzzy set have a continuum of membership grades ranging between zero and one (Zadeh, 1965).

Let C be a crisp universal set composed of *K* elements, and A a fuzzy subset of C. The value $\mu(A)_k$ represents the membership degree of element k $(k = 1, 2, ..., K)$ in A such that $0 \le \mu(A)_k \le 1$ and $0 \le \sum_{k=1}^{K} \mu(A)_k \le K$. In other words, $\mu(A)_k$ represents the degree of compatibility of the *k*-th element with the fuzzy set A (Klir and Wierman 1999). For further details on fuzzy set theory, see for example Dubois and Prade (1980), or Klir and Yuan (1995).

In community ecology, several authors have noted that functional resemblance among species can be framed within the context of fuzzy set theory. Specifically, the similarity between two species, *i* and *j*, represents the degree of membership (or functional compatibility) of individuals of one species to the set defined by the functional traits of the other species (Roberts 1986; Feoli and Zuccarello 1986; Feoli and Orlóci 2011; Ricotta and Pavoine 2025). Thus, by relaxing the mutually exclusive nature of distinct species, a higher similarity between species *i* and *j* implies a greater functional overlap between these two species (Figure 1). Note that, since by definition $s_{ii} = 1$, we assume that the membership degree of all individuals of species *i* to the fuzzy set defined by the functional traits of species *i* itself is always equal to one (Ricotta and Pavoine, 2025). Note also that a fuzzy set A for which there exists an element *k* such that $\max \mu(A)_k = 1$ is said to be normal. That is, a normal fuzzy set has at least one element with membership degree equal to one.

Unlike the interpretation of a species as a crisp set (in which an individual either fully belongs to the species or does not), a 'fuzzy species' represents a special case of a fuzzy set, which is defined by a core of characteristic or 'archetypal' elements (the individuals belonging to the focal species) together with several groups of 'satellite' individuals from different species that belong only partially to the set depending on their degree of functional similarity to the species under consideration.

Supposing that functionally similar species perform similar ecological roles, Leinster and Cobbold (2012) defined the ordinariness of species *i* as the commonness of all individuals in C that are functionally similar to *i* (including the individuals of species *i* itself):

$$\omega_i = \sum_{j=1}^{N} p_j s_{ij} \tag{13}$$

ω_i thus represents the abundance of all community individuals that support the functions associated with species *i*. If s_{ij} is bounded within the unit interval, we have $p_i \le \omega_i \le 1$. $\omega_i = p_i$ if all species $j \ne i$ are maximally dissimilar from *i*, while $\omega_i = 1$ if all species are functionally identical to *i* (i.e., if $s_{ij} = 1$ for all $j \ne i$).

By combining Eq. 12 and 13, it can be readily shown that, when intraspecific trait variability is neglected, quadratic diversity can be expressed as the expected functional rarity of species ordinariness ω_i (Ricotta and Szeidl 2006):

$$Q = \sum_{i=1}^{N} p_i (1 - \omega_i) \tag{14}$$

Therefore, Rao's *Q* can be regarded as a measure of conflict among functionally overlapping species. Since $\omega_i \ge p_i$, it follows that the value of Rao's index is always lower than the Simpson index: $S \ge Q$ where equality holds for a 'Boolean' community in which all species are considered maximally distinct from each other (i.e. with $d_{ij} = 1$ for all $j \ne i$, and $d_{ii} = 0$).

At least for ultrametric distances, Pavoine and Bonsall (2009) and Pavoine (2026) further showed that, for a fixed number of species and given pairwise distances, quadratic entropy is maximized when all species have equal functional rarity $1 - \omega_i$, mirroring the link between diversity and evenness proper to conventional abundance-based measures.

In terms of uncertainty, we recall that the term $\omega_i = \sum_{j=1}^{N} p_j s_{ij}$ represents the average degree of compatibility of the individuals in C with the functional characters of species *i*. Accordingly, the term $1 - \omega_i = \sum_{j=1}^{N} p_j d_{ij} = \sum_{j=1}^{N} p_j (1 - s_{ij})$ (i.e., the mean fuzzy complement of s_{ij} *sensu* Klir and Wierman 1999) expresses the average degree of *in*compatibility of the individuals in C with the functional characters of species *i*. At the same time, $1 - \omega_i$ can also be viewed as the abundance of all community individuals that conflict with the ecological functions performed by species *i*. Thus, in analogy with the interpretation of Simpson's index in a probabilistic context, Rao's quadratic diversity can be considered a measure of strife or discord for a set of fuzzy sets, such as that illustrated in Figure 1. That is, as a measure of uncertainty or disagreement in selecting among several partially overlapping alternatives.

Species-level functional distinctiveness and specificity

Functional distinctiveness or originality is a fundamental concept in ecology, describing the degree to which a species exhibits distinct functional traits compared to the other species within a community (Grenié et al. 2017; Violle et al. 2017). Although they may not be the most abundant species, those occupying extreme positions in trait space can make disproportionately large contributions to ecosystem processes. Such species often possess specialized traits that enable them to perform critical ecological roles, such as maintaining ecosystem functioning and stability (Delalandre et al. 2022; Munoz et al. 2023). Consequently, the loss of functionally unique species can trigger cascading effects, ultimately reducing ecosystem efficiency (Cadotte et al. 2011).

Unlike all community-level measures of diversity discussed so far, functional distinctiveness is therefore a species-level concept that measures the eccentric position of a species in trait space, with functionally distinct species being those located farther apart from other species (Violle et al. 2017; Mahaut et al. 2023). In terms of uncertainty, functional distinctiveness is grounded in the work of Yager (1982, 1992), who extended the notion of (non)specificity from crisp to fuzzy sets to summarize the extent to which a fuzzy set (a species) contains only a limited number of elements (individuals).

According to Yager (1982), for a normal fuzzy set A containing *K* discrete elements with varying membership degrees, a measure of specificity Sp in the range $[0,1]$ satisfies the following requirements: (1) Sp is maximum if there exists one and only one element k with $\mu(A)_k = 1$ and $\mu(A)_h = 0$ for all other elements *h* $(h \neq k)$; (2) Sp is minimum when $\mu(A)_k = 1$ for all $k \in K$; (3) For a normal fuzzy set with at least one element with $\mu(A)_k = 1$, any increase in the membership degree of the *h*-th element $(h \neq k)$ reduces (or at least does not increase) the specificity of the fuzzy set.

In this framework, since species ordinariness $\omega_i = \sum_{j=1}^{N} p_j s_{ij}$ summarizes the abundance of all community individuals that support the functions associated with species *i* (including the individuals of species *i* itself), its complement $1-\omega_i$, can be intuitively regarded as an appropriate index of fuzzy specificity for continuous abundance measures:

$$Sp_i = 1 - \omega_i = \sum_{j=1}^{N} p_j d_{ij} \tag{15}$$

Pavoine and Ricotta (2021) referred to this index as the 'effective originality' of species *i*. In Appendix 1, we show that this index can be directly derived from the work of Yager (1992). It follows that, if we consider $1-\omega_i$ as a measure of fuzzy specificity, Rao's quadratic diversity can be regarded as the average specificity of all species within the community (see Eq. 14). This perspective highlights the relationship between strife and mean functional specificity within a fuzzy representation of community structure.

This 'absolute' measure of specificity $1-\omega_i$ accounts for all individuals in C in determining the specificity of a focal species *i*. However, whereas in its classical definition, specificity quantifies the degree to which a fuzzy set contains only a single element (Klir and Wierman, 1999), the particular data structure of a fuzzy species leads us to interpret the specificity of species *i* as the degree to which the set contains only one *type* of elements. In this sense, the specificity of species *i* can also be regarded as the 'relative' contribution of individuals from other species in supporting the ecological role of species *i*.

When dealing with continuous abundance values, a common way to compute the (relative) distinctiveness of a focal species *i* is to compute the mean functional dissimilarity of *i* from all other species in the community (Ricotta et al. 2016; Violle et al. 2017):

$$Sp_i^* = \sum_{j=1, j\neq i}^{N} \frac{p_j}{1-p_i} d_{ij} \tag{16}$$

where $p_j/(1-p_i)$ denotes the abundance of species *j* excluding the focal species i $(j \neq i)$ such that $\sum_{j=1, j\neq i}^{N} p_j/(1-p_i) = 1$. In Appendix 1, we show how this index of specificity can also be derived from the work of Yager (1992).

Pavoine and Ricotta (2021) first noted that Eq. 16 can be expressed in terms of functional rarity as:

$$Sp_i^* = \frac{1-\omega_i}{1-p_i} \tag{17}$$

Therefore, relative functional distinctiveness Sp_i^* can be regarded as a scaled version of functional rarity, normalized by its local maximum $1-p_i$. That is, by the maximum value that $1-\omega_i$ can attain when the relative abundance of the focal species equals p_i. In essence, relative functional distinctiveness quantifies the extent to which the functional rarity of species *i* decreases when moving from a crisp model of mutually exclusive species to a fuzzy model of partially overlapping species in which each individual may simultaneously belong to several species to varying degrees.

If the functional dissimilarity d_{ij} between the focal species *i* and all other species *j* equals one, the fuzzy species will contain only one type of element (the individuals of species *i* itself), and its specificity will be maximal ($Sp_i^* = 1$). Conversely, if all species in C are functionally identical to species *i*, the fuzzy species will include all individuals in C with a membership degree equal to one. In this case, we obtain $Sp_i^* = 0$.

Ricotta and Pavoine (2025) observed that this way of looking at functional distinctiveness allows us to extend Eq. 16 to a general class of specificity measures expressed as the ratio between a species' functional rarity and its corresponding abundance rarity:

$$Sp_i^* = \frac{\rho(\omega_i)}{\rho(p_i)} \tag{18}$$

Since $p_i \leq \omega_i \leq 1$, we have $\rho(p_i) \geq \rho(\omega_i) \geq 0$, and therefore $0 \leq Sp_i^* \leq 1$.

As discussed by Ricotta and Pavoine (2025), absolute and relative specificity differ primarily in the perspective they represent. Absolute specificity Sp_i adopts the viewpoint of an external observer looking at all individuals in the community that can perform the functions attributed to a given species. In this sense, absolute specificity reflects the rarity of the species' functional characters, such as the state of a specific trait (Kondratyeva et al. 2019). In contrast, relative specificity Sp_i^* reflects the extent to which a species is functionally distinct from the other species in the community. Accordingly, when assessing relative specificity, the abundance of the focal species is not considered. Therefore, to convert relative specificity into its absolute counterpart, the relative measure needs to be multiplied by the corresponding abundance rarity: $Sp_i = Sp_i^* \times \rho(p_i)$.

From this perspective, it is also possible to define a general class of measures of fuzzy discord by generalizing Rao's formula as:

$$D = \sum_{i=1}^{N} p_i \times Sp(p_i) \tag{19}$$

This general formulation allows rare and abundant species to be weighted differently depending on the selected specificity index (Ricotta and Szeidl, 2006). In this way, for practical purposes, one is free to apply the type of rarity measure most appropriate for addressing the specific problem under consideration.

Discussion

By linking the most relevant diversity indices to their mathematical models of uncertainty, we aimed to provide a comprehensive theoretical framework that moves from a 'simplex' to a more 'complex' representation of diversity (in the sense of Juhász-Nagy 1993). In doing so, we hope to contribute to answering Juhász-Nagy's fundamental question: Do we know what we are talking about when we talk about diversity?

Historically, due to the lack of a clear biological meaning, traditional diversity indices received strong criticism from Hurlbert (1971), who regarded the Shannon index as a 'non-concept': "Although these information theoretic indices have been examined and applied to ecological problems by many ecologists, no one has yet specified exactly what [biological] significance the 'number of bits per individual' has to the individuals and populations in a community".

Hurlbert instead suggested adopting diversity measures based on species accumulation curves and rarefaction methods, which provide a more direct biological interpretation of ecosystem functioning. According to Hurlbert, this is because rarefaction methods are directly linked to intra- and interspecific encounters, which are in turn associated with several ecological interactions such as competition or predation. Interestingly, a few years later, Patil and Taillie (1982) used encounter theory as a general mathematical foundation to derive those diversity indices that Hurlbert intended to discard. Likewise, from the work of Chao and Jost (2012, Eq. 2 and 3) it is evident that the Simpson diversity is directly related to the slope of species accumulation curves, thereby challenging the presumed advantage of rarefaction methods as diversity measures.

In this context, while early debates on diversity were marked by an intense dualism between diversity as a biological property and diversity as a statistical index (discussed in Ricotta 2005), in this paper we adopt the view that diversity measures represent a class of multivariate summary statistics that condense various aspects of community structure into single scalars in a similar manner as similarity measures are used to quantify how closely two objects resemble each other (Ricotta 2005). Although similarity measures are not generally attributed intrinsic biological meaning, standard multivariate methods based on resemblance, such as ordination or clustering, have been of considerable importance for advancing our ecological understanding. Likewise, diversity measures have proven to be powerful tools for elucidating a wide range of ecological processes, including species co-occurrence and community assembly.

From a statistical viewpoint, the mathematical relationships among the different models of uncertainty explored in this work allowed us to define a new class of measures of strife or discord for a set of fuzzy sets, and to express these measures in terms of expected specificity. Although the concept of specificity was introduced in fuzzy set theory by Yager as early as 1982, in ecology the relationship between specificity and functional originality has been highlighted only recently by Ricotta and Pavoine (2025). In this work, we have addressed several open issues left by Ricotta and Pavoine (2025), thus providing a comprehensive theoretical framework that allows us to express a standard measure of functional diversity, such as Rao's Q, in terms of the mean functional distinctiveness of all species in the community. This has been made possible by defining a general class of specificity measures directly derived from Yager's (1992) work, which can also be extended to continuous measures of quantity. Indeed, the observation that the calculation of a diversity index cannot be restricted to discrete counts of individuals dates back to early work on biodiversity (Hill 1973; Tóthmérész 1995). For instance, Hill (1973) notes that in vegetation science counts of individual plants are usually impossible. Therefore, we need to relax the strict individual-based interpretation of diversity measures and compute community diversity in terms of continuous values of species abundance.

We also introduced a distinction between absolute and relative indices of functional specificity, which is not present in the classical literature on fuzzy specificity (e.g., Yager 1982, 1992; Klir and Wierman 1999; Marín et al. 2020). Absolute indices refer to the abundance of all community individuals that support the functions associated with a focal species. Therefore, the mean value of these indices, computed across all species in the assemblage represents a measure of functional diversity suitable for describing community functioning (see Eq. 19). On the other hand, relative specificity reflects the degree of functional differentiation of a focal species from the other community species, without taking its own abundance into account (Scheiner 2019). As such, relative specificity provides an appropriate species-level indicator of functional originality to be used, for example, in conservation actions aimed at identifying functionally unique species that play irreplaceable roles in ecosystem functioning (Griffith et al. 2021; Ricotta et al. 2025).

With regard to the measurement of strife, to the best of our knowledge this is the first time that a measure of strife has been proposed for a set of fuzzy sets and that such a measure is expressed in terms of expected specificity. This makes the two types of uncertainty far more closely interconnected than is generally assumed. Indeed, the common view is that measures of strife, such as the Shannon entropy, are typically associated with a probabilistic model of uncertainty, whereas specificity measures, such as Hartley's formula, summarize uncertainty in classical set theory, of which fuzzy set theory is a generalization. According to Klir and Wierman (1999): "We recognize now that the Hartley function and the Shannon entropy measure distinct types of uncertainty, nonspecificity and strife, respectively". The view that the Hartley function is a special case of the Shannon entropy which stems from the observation that the value of the Shannon entropy for a uniform probability distribution equals the value of the Hartley function was, in their words, "ill-conceived".

However, although Klir and Wierman correctly argue that strife and specificity generally represent two distinct types of uncertainty, the Shannon index is nevertheless directly derived from Hartley's formula (see above). Moreover, in the particular case of a perfectly even probability distribution, $p_1 = p_2 = ... = p_N = 1/N$ the two uncertainty models collapse into one another. This is because for an equiprobable distribution the individual probabilities, being all equal, essentially lose their informative role, such that no meaningful distinction remains between specificity and strife. In this case, since there is no substantial difference between the two types of uncertainty, the Shannon index reduces to Hartley's formula. More generally, a relevant distinction between the two types of uncertainty is that Hartley's formula quantifies the uncertainty associated with a single crisp set, whereas the Shannon index measures the uncertainty associated with a set of crisp (sub)sets of distinct events (see Figure 1). Accordingly, we can say that specificity and strife quantify uncertainty *within* and *among* sets (whether crisp or fuzzy), respectively.

Another relevant aspect of the relationship between Rao's quadratic diversity and strife is that, while the Rao index offers ample (and desirable) flexibility in the choice of the dissimilarity coefficient (Ricotta and Moretti 2011), if the functional dissimilarity between two species is computed from a single (univariate) trait τ using half the squared Euclidean distance $d_{ij} = \frac{1}{2}\left(\tau_i - \tau_j\right)^2$, quadratic diversity equals the variance of τ (Rao 2010; Pavoine 2012). Therefore, quadratic diversity is usually regarded as a multivariate analogue of variance.

From an ecological viewpoint, since functional diversity is essentially related to trait variation or dispersion within communities (Gregorius and Kosman 2017), it is not surprising that Rao's index can be effectively used for assessing functional diversity. What is more surprising, however, is that a measure of strife for a set of fuzzy sets can be expressed in terms of a multivariate analogue of variance, a standard tool of classical statistics.

To conclude, relating the most commonly used diversity indices to their underlying models of uncertainty allowed us to examine both aspects from a non-standard perspective, leading to a productive cross-fertilization between the two research fields. From the standpoint of community ecology, it enabled us to express Rao's functional diversity as the mean absolute originality of the distinct species in a given assemblage; from the standpoint of fuzzy set theory it allowed us to define a new class of measures of strife and to make significant progress in the definition and understanding of specificity indices. Taken together, these results offer a more integrated perspective on the two approaches, which we hope will inspire further advances at the intersection of ecological diversity and uncertainty theory.

References

Alroy, J. (2025) Does evenness even exist? Ecology Letters 28: e70181.

Cadotte, M.W., Carscadden, K., Mirotchnick, N. (2011) Beyond species: Functional diversity and the maintenance of ecological processes and services. Journal of Applied Ecology 48: 1079–1087.

Champely, S., Chessel, D. (2002) Measuring biological diversity using Euclidean metrics. Environmental and Ecological Statistics 9: 167–177.

Chao, A., Chiu, C.-H., Jost, L. (2014) Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. Annual Review of Ecology, Evolution and Systematics 45: 297–324.

Chao, A., Jost, L. (2012) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. Ecology 93: 2533–2547.

Delalandre, L., Gaüzère, P., Thuiller, W., Cadotte, M., Mouquet, N., Mouillot, D., Munoz, F., Denelle, P., Loiseau, N., Morin, X., Violle, C. (2022) Functionally distinct tree species support long-term productivity in extreme environments. Proceedings of the Royal Society B-Biological Sciences: 289: 20211694.

Díaz, S., Cabido, M., 2001. Vive la difference: plant functional diversity matters to ecosystem processes. Trends in Ecology and Evolution 16: 646–655.

Dubois, D., Prade, H. (1980) Fuzzy sets and systems - Theory and applications. Academic Press, New York.

Feoli, E., Orlóci, L. (2011) Can similarity theory contribute to the development of a general theory of the plant community? Community Ecology 12: 135–141.

Feoli, E., Zuccarello, V. (1986) Ordination based on classification: yet another solution?! Abstracta Botanica 10: 203–219.

Fisher, R.A., Corbet, A.S., Williams, C.B. (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. Journal of Animal Ecology 12: 42–58.

Good, I.J. (1953) The population frequencies of species and the estimation of population parameters. Biometrika 40: 237–264.

Gregorius, H.R., Kosman, E. (2017) On the notion of dispersion: From dispersion to diversity. Methods in Ecology and Evolution 8: 278–287.

Grenié, M., Denelle, P., Tucker, C.M., Munoz, F., Violle, C. (2017) funrar: an R package to characterize functional rarity. Diversity and Distributions 23: 1365–1371.

Griffith, P., Lang, J.W., Turvey, S.T., Gumbs, R. (2021) Using functional traits to identify conservation priorities for the world's crocodylians. Functional Ecology 37: 112–124.

Hendrickson, J.A.J., Ehrlich, P.R. (1971) An expected concept of "species diversity". Notulae Naturae 439: 1-6.

Hill, M. (1973) Diversity and evenness: a unifying notation and its consequences. Ecology 54: 427–432.

Hurlbert, S.H., (1971) The nonconcept of species diversity: a critique and alternative parameters. Ecology 52: 577–586.

Izsák, J., Papp, L. (1995) Application of the quadratic entropy index for diversity studies on drosophilid species assemblages. Environmental and Ecological Statistics 2: 213-224.

Juhász-Nagy, P. (1993) Notes on compositional diversity. Hydrobiologia 249: 173-182.

Juhász-Nagy, P., Podani, J. (1983) Information theory methods for the study of spatial processes and succession. Vegetatio 51: 129-140.

Klir, G.J., Wierman, M.J. (1998) Uncertainty-Based Information: Elements of Generalized Information Theory. Physica-Verlag, New York.

Klir, G.J., Yuan, B. (1995) Fuzzy Sets and Fuzzy Logic. Theory and Applications. Prentice-Hall, Upper Saddle River, NJ.

Kondratyeva, A., Grandcolas, P., Pavoine, S. (2019) Reconciling the concepts and measures of diversity, rarity and originality in ecology and evolution. Biological Reviews 94: 1317–1337.

Laliberté, E., Legendre, P. (2010) A distance-based framework for measuring functional diversity from multiple traits. Ecology 91: 299–305.

Leinster, T., Cobbold, C.A. (2012) Measuring diversity: the importance of species similarity. Ecology 93: 477–489.

MacArthur, R.H. (1955) Fluctuation of animal populations and a measure of community stability. Ecology 36: 533–536.

MacArthur, R., MacArthur, J.W. (1961) On Bird Species Diversity. Ecology 42: 594–598.

Mahaut, L., Violle, C., Shihan, A., Pélissier, R., Morel, J.-B., de Tombeur, F., Rahajaharilaza, K., Fabre, D., Luquet, D., Hartley, S., Thorne, S. J., Ballini, E., Fort, F. (2023) Beyond trait distances: Functional distinctiveness captures the outcome of plant competition. Functional Ecology 37: 2399–2412.

Margalef, D.R. (1958) Information theory in ecology. International Journal of General Systems 3: 36–71.

Marín, N., Rivas-Gervilla, G., Sánchez, D., Yager, R.R. (2020) Specificity measures based on fuzzy set similarity. Fuzzy Sets and Systems 401: 189–199.

Molinari, J. (1989) A calibrated index for the measurement of evenness. Oikos 56: 319–326.

Mouchet, M.A., Villéger, S., Mason, N.W.H., Mouillot, D. (2010) Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. Functional Ecology 24: 867–876.

Munoz, F., Klausmeier, C. A., Gaüzère, P., Kandlikar, G., Litchman, E., Mouquet, N., Ostling, A., Thuiller, W., Algar, A. C., Auber, A., Cadotte, M. W., Delalandre, L., Denelle, P., Enquist, B.J., Fortunel, C., Grenié, M., Loiseau, N., Mahaut, L., Maire, A., Mouillot, D., Pimiento, C., Violle, C., Kraft, N.J.B. (2023) The ecological causes of functional distinctiveness in communities. Ecology Letters 26: 1452–1465.

Patil, G.P., Taillie, C. (1982) Diversity as a concept and its measurement. Journal of the American Statistical Association 77: 548–561.

Pavoine, S. (2012) Clarifying and developing analyses of biodiversity: towards a generalisation of current approaches. Methods in Ecology and Evolution 3: 509–518.

Pavoine, S. (2026) The Rich-Rao Quadratic Entropy. Ecological Indicators 182: 114512.

Pavoine, S., Bonsall, M.B. (2009) Biological diversity: distinct distributions can lead to the maximization of Rao's quadratic entropy. Theoretical Population Biology 75: 153–163.

Pavoine, S., Ricotta, C. (2021) On the relationships between rarity, uniqueness, distinctiveness, originality and functional/phylogenetic diversity. Biological Conservation 263: 109356.

Peet, R.K. (1974) The measurement of species diversity. Annual Review of Ecology and Systematics 5: 285–307.

Podani, J. (2000) Introduction to the Exploration of Multivariate Biological Data. Backhuys, Leiden, NL.

Rao, C.P. 1982. Diversity and similarity coefficients: a unified approach. Theoretical Population Biology 21: 24–43.

Rao, C.R. (2010) Quadratic entropy and analysis of diversity. Sankhya 72: 70–80.

Rényi, A. (1970) Probability theory. North-Holland Publishing, Amsterdam, NL.

Ricotta, C. (2005) Through the jungle of biological diversity. Acta Biotheoretica 53: 29–38.

Ricotta, C., Bacaro G., Maccherini S., Pavoine S. (2022) Functional imbalance not functional evenness is the third component of community structure. Ecological Indicators 140: 109035.

Ricotta, C., Dalle Fratte, M., Cerabolini, B.E.L., Pavoine, S. (2025) Beyond mean trait distance: an integrated approach to functional distinctiveness. Ecological Indicators 181: 114477.

Ricotta, C., de Bello, F., Moretti, M., Caccianiga, M., Cerabolini, B., Pavoine, S. (2016) Measuring the functional redundancy of biological communities: a quantitative guide, Methods in Ecology and Evolution 7: 1386-1395.

Ricotta, C., Feoli, E. (2024) Hill numbers everywhere. Does it make ecological sense? Ecological Indicators 161: 111971.

Ricotta, C., Kosman, E., Laroche, F., Pavoine, S. (2021) Beta redundancy for functional ecology. Methods in Ecology and Evolution 12: 1062–1069.

Ricotta, C., Moretti, M. (2011) CWMand Rao's quadratic diversity: a unified framework for functional ecology. Oecologia 167: 181–188.

Ricotta, C., Pavoine, S. (2025) What do functional diversity, redundancy, rarity, and originality actually measure? A theoretical guide for ecologists and conservationists. Ecological Complexity 61: 101116.

Ricotta, C., Podani, J., Schmera, D., Bacaro, G., Maccherini, S., Pavoine, S. (2023). The ternary diagram of functional diversity. Methods in Ecology and Evolution 14: 1168–1174.

Ricotta, C., Szeidl, L. (2006) Towards a unifying approach to diversity measures: Bridging the gap between the Shannon entropy and Rao's quadratic index. Theoretical Population Biology 70: 237–243.

Roberts, D.W. (1986) Ordination on the basis of fuzzy set theory. Vegetatio 66: 123–143.

Scheiner, S.M. (2019) A compilation of and typology for abundance-, phylogenetic- and functional-based diversity metrics. bioRxiv. https: //doi.org/10.1101/530782.

Schmera, D., Erös, T., Podani, J. (2009) A measure for assessing functional diversity in ecological communities. Aquatic Ecology 43: 157–167.

Shannon, C. (1948) A mathematical theory of communication. Bell System Technical Journal 27: 379–423.

Sherwin, W.B., Jabot, F., Rush, R., Rossetto, M. (2006) Measurement of biological information with applications from genes to landscapes. Molecular Ecology 15: 2857– 2869.

Shimatani, K. (2001) On the measurement of species diversity incorporating species differences. Oikos 93: 135–147.

Simpson, E.H. (1949) Measurement of diversity. Nature 163: 688.

Solow, A.R., Polasky, S. (1994) Measuring biological diversity. Environmental and Ecological Statistics 1: 95–103.

Villéger, S., Mason, N.W.H., Mouillot, D. (2008) New multidimensional functional diversity indices for a multifaceted framework in functional ecology. Ecology 89: 2290–22301.

Violle, C., Thuiller, W., Mouquet, N., Munoz, F., Kraft, N.J.B., Cadotte, M.W., Livingstone, S.W., Mouillot, D. (2017) Functional rarity: The ecology of outliers. Trends in Ecology and Evolution 32: 356–367.

Walker, B., Kinzig, A., Langridge, J. (1999) Plant attribute diversity, resilience, and ecosystem function: the nature and significance of dominant and minor species. Ecosystems 2: 95–113.

Yager, R.R. (1982) Measuring tranquility and anxiety in decision making: An application of fuzzy sets. International Journal of General Systems 8: 139–146.

Yager, R.R. (1992) Default knowledge and measures of specificity. Information Sciences 61: 1–44.

Zadeh, L. (1965) Fuzzy sets. Information and Control 8: 338–353.

Funding information

CR was supported by a research grant from the University of Rome ‘La Sapienza’ (RM124190C044444C).

CRediT authorship contribution statement

CR: Conceptualization, Methodology, Formal analysis, Writing – original draft; AC: Methodology, Formal analysis, Writing – review & editing; SP: Methodology, Formal analysis, Writing – review & editing; DC: Methodology, Formal analysis, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

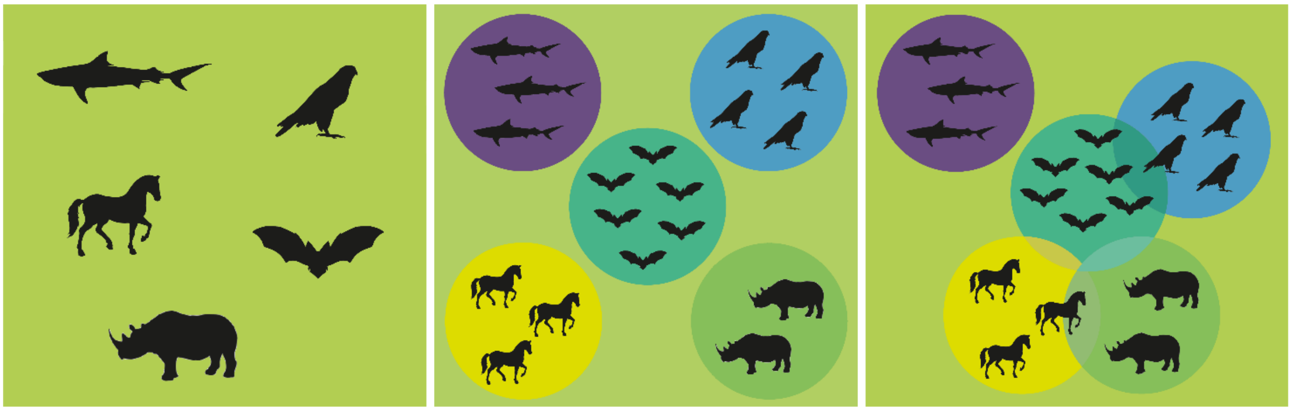

Figure 1. Schematic representation of the different models of uncertainty-based diversity discussed in this paper. **Left – Classical set theory:** the community is represented as a crisp set containing a finite collection of distinct elements (species). Here, the emphasis is on the number of distinct species, regardless of their abundances. **Center – Probabilistic uncertainty:** the community is represented as a set of crisp subsets (species), each containing a certain number of individuals. Each individual cannot simultaneously belong to two or more species. Therefore, species are considered mutually exclusive. **Right – Fuzzy uncertainty:** the community is represented as a set of partially overlapping fuzzy subsets (species). Unlike the probabilistic case, individuals may simultaneously belong to several species to different degrees according to their functional (or taxonomic) similarity.

Appendix 1. Quantifying the specificity of fuzzy species

Yager (1982) extended the concept of (non)specificity from crisp sets to fuzzy sets to quantify the degree to which a fuzzy set contains only one element (Klir and Wierman 1999).

Let C be a crisp set composed of *K* elements, and let A be a fuzzy subset on C where the value $\mu(A)_k$ (hereafter denoted simply as μ_k) represents the membership degree of element k $(k = 1, 2, ..., K)$ in A with $0 \leq \mu_k \leq 1$. That is to say, μ_k expresses the degree of compatibility of the *k*-th element with the concept represented by the fuzzy set A. For simplicity, we assume that the values of μ_k are ranked such that $\mu_k \geq \mu_{k+1}$.

According to Yager (1982), for a normal fuzzy set A containing *K* discrete elements, a measure of specificity *Sp* in the range $[0,1]$ should satisfy the following basic requirements (recall that a fuzzy set is said to be normal if it contains at least one element with a membership degree $\mu_k = 1$):

R1. $Sp_A = 1$ if there exists one and only one element *k* with membership degree $\mu_k = 1$ and $\mu_h = 0$ for all other elements h $(h \neq k)$

R2. Sp_A is minimum if $\mu_k = 1$ for all $k \in K$

R3. For a normal fuzzy set with at least one element $\mu_k = 1$, any increase in the membership degree of the *h*-th element $(h \neq k)$ decreases (or at least does not increase) the specificity of A

Based on requirements R1–R3, Yager (1992) introduced a family of so-called ‘linear’ specificity measures, expressed as:

$$Sp_A = w_k \times \mu_k - \sum_{h=2}^{K} w_h \times \mu_h \tag{1}$$

where the weights *w* satisfy the following conditions (Yager 1992):

C1. $w_k = 1$

C2. $0 \leq w_h \leq 1$, $(h \neq k)$

C3. $\sum_{h=2}^{K} w_h \leq 1$

C4. $w_h \geq w_{h+1}$ for a ranked distribution $\mu_h \geq \mu_{h+1}$

According to Yager (1992), the idea behind condition C4 is that the weights are not associated with particular elements but rather with ordered positions.

In the specific case of a biological community, let C be a crisp set (the community) composed of *K* individuals, and let K_i be the number of individuals belonging to species i $(i = 1, 2, ..., N)$ such that $p_i = K_i / K$ denotes the relative abundance of the individuals of species *i*. As in the main text, we assume that the membership degree of the individuals of species *j* to the normal fuzzy set defined by species *i* corresponds to the functional similarity between *i* and *j*.

If the membership degrees to species *i* of the *K* individuals are ranked such that $s_{ik} \geq s_{ih} \geq s_{ih+1}$ with $h \neq k$ and $s_{ik} = 1$, and the weights w_h $(h = 2,...,K)$ are all equal to $1/(K-1)$, according to Eq. 1, the specificity of the fuzzy species *i* can be expressed as:

$$Sp_i = w_k \times s_{ik} - \sum_{h=2}^{K} w_h \times s_{ih} = 1 - \frac{1}{K-1} \sum_{h=2}^{K} s_{ih} \tag{2}$$

which is simply the difference between the maximum membership degree to species *i* of the *K* individuals in C, and the mean of the membership values of the remaining $K-1$ individuals.

By grouping individuals by species, Eq. 2 can be rewritten as:

$$Sp_i = 1 - \left(\frac{K_i - 1}{K-1} \times s_{ii} + \sum_{j=2}^{N} \frac{K_j}{K-1} \times s_{ij} \right) \tag{3}$$

If the abundances across species are held constant, for sufficiently large *K* we have $(K_i - 1)/(K-1) \approx K_i/K$ and $K_j/(K-1) \approx K_j/K$, and hence:

$$Sp_i \approx 1 - \sum_{j=1}^{N} \frac{K_j}{K} \times s_{ij} = 1 - \sum_{j=1}^{N} p_j \times s_{ij} = 1 - \omega \tag{4}$$

Thus recovering the formula of absolute specificity for continuous abundance values given in Eq. 15 of the main text.

According to this classical way of looking at specificity, the functional distinctiveness of focal species *i* increases both with decreasing abundance of the individuals of species *i*, and with increasing functional similarity between the focal species and the other species in the community.

However, beyond this conventional definition of specificity, the data structure of a fuzzy species leads us to interpret the specificity of species *i* as the degree to which the species contains only one *type* of element.

To satisfy this aspect of specificity, it is convenient to define the fuzzy set over the *N* species in the community rather than over the *K* individuals. In this case, Yager's requirements R1–R3 for a normal fuzzy species with $s_{ii} = 1$ become:

R1'. For a community composed of *N* species, the specificity of focal species *i* is maximum $(Sp_i^* = 1)$ if $s_{ij} = 0$ for all species *j* $(j \neq i)$.

R2'. Sp_i^* is minimum if $s_{ij} = 1$ for all species *j* $(j \neq i)$.

R3'. Any increase in the membership degree of species *j* to the focal species *i* $(j \neq i)$ decreases (or at least does not increase) the specificity of *i*.

By relaxing condition C4 and associating the weights *w* with the species in C rather than with ordered positions, the linear measure in Eq. 1 can be expressed over the *N* species as:

$$Sp_i^* = w_i \times s_{ii} - \sum_{j=2}^{N} w_j \times s_{ij} = 1 - \sum_{j=2}^{N} \frac{p_j}{1-p_i} \times s_{ij} = \sum_{j=2}^{N} \frac{p_j}{1-p_i} \times d_{ij} \tag{5}$$

where $w_i = 1$ and $w_j = p_j / (1-p_i)$. In this case, the weights w_j represent the scaled relative abundances of species *j* when the focal species *i* is not considered.

It can be easily shown that Eq. 5 can be expressed in terms of functional rarity as (Pavoine and Ricotta 2021; Ricotta and Pavoine 2025):

$$Sp_i^* = \frac{1-\omega_i}{1-p_i} \tag{6}$$

By relaxing condition C1, we can further generalize Eq. 6 to define a general class of relative specificity measures which recovers Eq. (18) of the main text:

$$Sp_i^* = \frac{\rho\left(\sum_{j=1}^{N} w_j \times s_{ij}\right)}{\rho\left(w_i \times s_{ii}\right)} = \frac{\rho(\omega_i)}{\rho(p_i)} \tag{7}$$

with $w_i = p_i$ and $w_j = p_j$.

Note that, although Eq. 5-7 do not comply with conditions C1 and C4, they still conform to the conditions $\mathrm{R1'} - \mathrm{R3'}$ required for relative measures of specificity.

References

Klir, G.J., Wierman, M.J. (1999) Uncertainty-Based Information: Elements of Generalized Information Theory. Physica-Verlag, New York.

Pavoine, S., Ricotta, C. (2021) On the relationships between rarity, uniqueness, distinctiveness, originality and functional/phylogenetic diversity. Biological Conservation 263: 109356.

Ricotta, C., Pavoine, S. (2025) What do functional diversity, redundancy, rarity, and originality actually measure? A theoretical guide for ecologists and conservationists. Ecological Complexity 61: 101116.

Yager, R.R. (1982) Measuring tranquility and anxiety in decision making: An application of fuzzy sets. International Journal of General Systems 8: 139–146.

Yager, R.R. (1992) Default Knowledge and Measures of Specificity. Information Sciences 61: 1–44.