Index Systems: Enumerating Their Forms and Explaining Their Diversity With Representational Interpretive Structure Theory

Peter C-H. Cheng (p.c.h.cheng@sussex.ac.uk), Grecia Garcia Garcia (g.garcia-garcia)
Department of Informatics, University of Sussex, Brighton, BN1 9QJ, UK.

Daniel Raggi (dr495.@cam.ac.uk), Mateja Jamnik (mateja.jamnik@cl.cam.ac.uk)
Department of Computer Science and Technology, University of Cambridge, UK.

Abstract

Index systems are central to our everyday and intellectual lives. Their ubiquity and diversity make them an important class of cognitive artifacts, the study of which has implications for our understanding of representational systems in general. This paper builds schema-theoretic network models of the nature of the memory structures, that underpin the interpretation of indexing systems. We identify four common classes of index systems. Using *Representation Interpretation Structure Theory*, we explain how the four basic classes can be responsible for the substantial diversity among index systems.

Keywords: representational systems; index systems; interpretation; schema networks; numeration systems; cognitive artefacts

Introduction

We inhabit informational environments that are richly populated with diverse species of representational systems. Research investigating how representational systems support thinking and learning is well establish (e.g., Glasgow, Narayanan & Chandrasekaran, 1995; the Diagrams Conference series, e.g., Giardino et al., 2022). Work on reasoning with particular classes of representations (e.g., Cleveland, 1985; Shah, Mayer & Hegarty, 1999) and general properties of representations (Kirsh, 2010; Shimojima, 2015) are common foci for such research.

The focus here is a on a class of cognitive artifacts that have been rather neglected by studies of representational systems – *indexing systems*. Just a little reflection reveals that they are ubiquitous and suffuse our daily and intellectual lives. Table 1 presents a small selection of index systems. (The four-way classification is explained in Section 2.) Innumerable others could have been included, not in the least because they are legion in specialist technical domains. They are certainly diverse. The general function of index systems is to support the processing of information about *instances* (individuals, cases, datapoints, etc.). They are used to identify, characterize, arrange and retrieve instances. Although the cognition of some index systems has been examined, including numeration systems (Zhang & Norman, 1994b; Chrisomalis, 2020) and book subject indexes (Collison, 1972; Farrow, 1991), it seems that a general systematic account of the cognitive nature of indexing systems in general is yet to be given. So, there are two obvious questions that should be addressed in preliminary studies of the general nature of index systems.

(1) *What are the common classes of index systems, in general?* It appears likely that a small number of basic forms of indexing systems exist. We likely acquire them early during education, because we (adults) are readily able to adopt index systems that are novel to us without substantial training, and often with no explicit explanation. For instance, when we first use a new catalogue, website, appliance or other source of rich no-verbal information, we do not expect to find a special explanatory page on the to-be-encountered *index systems*. As anticipated by the organization of Table 1, our claim is that there are four such underlying classes: *metric indexes*, *taxonomic codes*, *triangulational coordinate systems,* and *mereological coordinate systems*. These will be introduced in Section 2.

(2) *If there are just a few basic forms, how should we explain the substantial diversity of specific types of index systems?* This question will be addressed by building theoretical models to show how each of the basic forms provides a space of possible designs.

Theoretical modelling building is an established approach to the study of the nature of representational systems in reasoning and problem solving in cognitive science. Various methods are common, including: computational modelling (e.g., Larkin & Simon, 1978; Tabachneck-Schijf, Leonardo & Simon, 1997; Cheng, 1996; Peebles & Cheng, 2003; Kunda, McGreggor, & Goel, 2013); diagramming network models (e.g., Pinker, 1990); formulating dimensional frameworks (e.g., Zhang & Norman,1994a, 1994b; Zhang, 1996); and, of course, expressing models verbally (e.g., Carpenter & Shah,1998).

To examine the diversity of index systems we adopt *Representational Interpretive Structure Theory*, *RIST*, (Cheng, 2020; Cheng, Stockdill, Garcia Garcia, Raggi & Jamnik, 2022; Stockdill, Garcia Garcia, Cheng, Raggi & Jamnik, 2022; Cheng, Garcia Garcia, Raggi & Jamnik, 2024). RIST is selected for three reasons. First, it is a schema-based theory that allows models of memory structures to be specified according to a set of theoretical assumptions tailored to representational systems. Second, it is accompanied by a graphical notation (*RIS-Notation*, *RISN)* that provides a rigorous operationalization of the theory, so the models conform to the theoretical assumptions of RIST. Third, the approach includes a method for building models of interpretations in a fully featured web-browser tool (*RISE*) for constructing and *E*diting RISN models. Section 3 summarises RIST, RISN

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Table 1. Four basic forms of index systems.

(a) Metric indexes[1]	**(b) Taxonomic codes**
N-N: Croquet set – colours of ball match to colour of mallets. N-O: Alphabetic lists; computer program line numbers; letters indexing this list (explained below); book subject indexes (Collison, 1972). N-I: Student registration numbers. O-O: Assignments grades (A+ to F); UK degree classes (1st, 2.1, 2.2, 3rd, Pass); book content pages. O-N: Manual car gearstick/stick shift. O-I: Numbered sections of this paper. I-I: Student assignment % marks; thermometer scale. R-R: Pie chart; dial timers; tape measures and rulers; water faucet/tap (rotary); gas/accelerator pedal. I-I*2+: Ancient Greek numeration systems (e.g., $\sigma\kappa\beta = 222$) (Zhang & Norman, 1994b); sum of two or more dice.[2]	• Zoological trinomial nomenclature: genus, species, subspecies; e.g., Eastern low land gorilla – *Gorilla beringei graueri*; Cross river gorilla – *Gorilla gorilla diehli*; Western lowland gorilla – *Gorilla gorilla gorilla*. • Academic subject codes, e.g.: Library of Congress and Dewey Decimal Classifications; APA PsycInfo Classification Categories and Codes (e.g., *2300 Human expt. psychology* contains *2340 Cognitive processing*). • Computer navigation: breadcrumb trails (e.g.; “Home>Contacts>Alex”); *column* format in macOS *Finder*. • Traditional Chinese names, comprising *family*, *generation* and *given* names (Kałużyńska, 2015).
(c) Triangulational coordinate systems[3]	**(d) Mereological coordinate systems**[3]
• Line graphs (2+); bar charts (2+), tables (2+); maps globes, astronomical planispheres (2); traffic lights (3) • Chess, checkers/drafts and Go boards (2); crossword puzzle grid (2); playing cards (3). • Airplane cockpit joystick (2); WIMP-style interface window scrollbars (2); domestic mixer faucet/tap (2); foot pedals of a manual/stick shift car (3). • Car registration plates (UK: location, year, item) (3); ISBNs (4); the list of references below (4+).	• Hindu-Arabic number place value (>1). • Hindu-Arabic numeration system overall structure (≥2) (Zhang & Norman, 1994b, Chrisomalis, 2020). • Multi-part alphanumeric lists (>1). • Clocks and dates (≥2). • Office room numbers (e.g., building, floor, room) (3). Street addresses (≥3), ZIP codes (3), UK Postal codes (4, typically).

[1] The letter pair codes are quantity scales of the concept and graphic object: Nominal, Ordinal, Interval, Ratio.
[2] I-I*2+ means the one interval scale for the concept is mapped to two or more interval scale graphic objects.
[3] The number in parentheses is the number of properties or R-dimensions.

and RISE. Following it, Section 4 presents and examines the models for each of the four classes of indexing systems.

Index systems

The first question of this paper asks: what are the basic underlying forms of index systems? We propose four forms: (a) *metric indexes*, (b) *taxonomic codes*, (c) *mereological coordinate* systems, and (d) *triangulational coordinate systems*. Examples of each of the systems a given in Table 1.

Diagrams of the informational structure of the four types of indexing systems are shown in Figure 1. The labels in the leaf node boxes are instances that are indexed. The intermediate nodes are properties, classes or values that distinguish instances. Paths from the domain to the leaves encode the structure of indexes.

Metric indexes are the simplest. They use a single property to differentiate a class of instances. In Figure 1a, node *N* stands for the property. Examples are given in Table 1a.

Instances in *taxonomic codes* are characterized by a succession of properties that are organized in levels within a tree-like structure, as depicted in Figure 1b. Importantly for this indexing system, the properties at the same level in the tree include distinct sub-properties for each superordinate element (in Figure 1b, Arabic versus Roman numerals). In other words, each level includes disjoint subclasses and the paths down the tree do not overlap. A path from the root to a leaf gives the code for an instance corresponding to the leaf. Examples are given in Table 1b.

Triangulational coordinate systems are defined by at least two properties that all contribute values to instances by a process of *triangulation* (Figure 1c). This class of indexes is similar to the mereological system, but its properties do not have a consistent order conceptually. It was previously identified by Stockdill et al. (2022) and called *coordinate systems*, and they defined two types: *explicit* and *implicit* (triangulational) coordinate systems. In an explicit system one

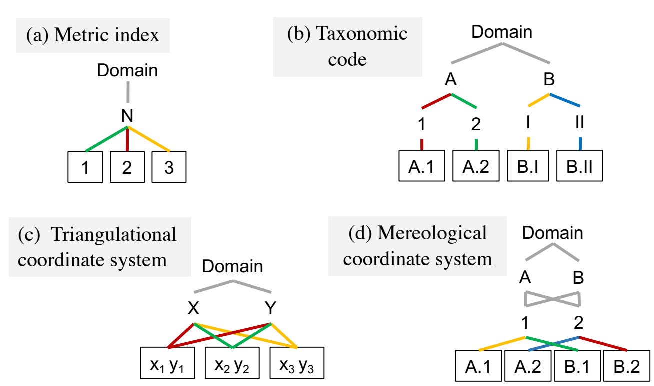

Figure 1. Informational structure of the indexing systems.

property specifically identifies a set of instances to interpret as datapoints whose other property values are to be determined by the coordinate system. Examples are given in Table 1c.

Mereological coordinate systems are defined by: (a) an ordering of properties across levels; (b) elements at one level are considered as parts of the element at its superordinate level; (c) at each level there is a single shared sub-property or class, as depicted by the shared (grey) paths in Figure 1d. They differ from taxonomic codes in that elements are constituent parts, rather than being members of a class, and at each level they are a property of entities, rather than classes of disjoint entities. Examples are given in Table 1d.

The four classes of index systems are clearly distinct. No claim is made that the four are exhaustive, but from the examples in Table 1, it is clear that the scope of the four is substantial. The examples also emphasise the diversity of the different types of systems within each of the classes. That leads us to the second question of how to explain the diversity. But first, Representation Interpretive Structure Theory is summarised.

Theory, Notation and Editor

Representational Interpretive Structure Theory, RIST, was proposed by Cheng (2020) and colleagues (Cheng et al., 2022, 2024; Stockdill et al., 2022; Cheng at al., 2024) as an approach to understanding the interpretation of representations and representational systems. It comprises four core ideas.

(1) There is a set of elementary (atomic) memory components, schemas, that encode the information associated with an interpretation. All the schemas (i) store information about the domain concept being represented and (ii) hold codes for the *graphic object* that represents the concept. A graphic object is any graphic entity (visuospatial properties and relations, icon, glyphs, etc.). In contrast with other accounts of the nature of representations that focus upon internal mental components of representational systems (Palmer, 1978; Kosslyn, 1989; Pinker, 1990), RIST claims that the primary function of the schemas deployed for the mental process of interpreting a representation is to tie information about the concepts being represented to information about the graphic objects in the external representation that are doing the representing (Cheng, 2002). RIST gives equal status to mental concepts and graphical objects and so contrasts with Zhang & Norman's (1994a, 1996) account of the distributed nature of representations, which conceptualizes the relation between concepts and graphic objects (symbols) hierarchically, with the external graphic objects as leaves.

(2) There are four core types of schemas: *Representation*, *R-scheme*, *R-dimension* and *R-symbol*. A graphical notation, *RISNotation* (RISN), for building models in RIST has been developed (Cheng, 2020). To introduce the schemas and RISN, Figure 2 shows a RISN model for the naming convention of Airbus airplanes, which is a composite indexing system. *Representation* schemas – **capsule** shape – define a domain of interest (Airbus airplanes) and identify the display

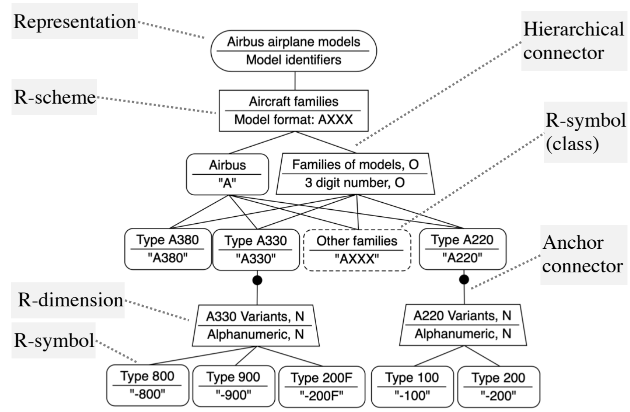

Figure 2. RISN model components.

for the domain (airplane names). *R-symbols* – **rounded rectangles** – encode fixed value concepts and the graphic objects representing them. RISN models have R-symbols as leaf nodes, in our example they are instances (airplane types; e.g., *A380*, *A330-800*). The idea of the manufacturer, *Airbus*, is also an R-symbol with "A" as its graphic object. There are also R-symbols for particular families of airplanes, which may or may not also be a type of an airplane. *Class* R-symbols – **dashed rounded rectangles** – are used to represent multiple R-symbols that are not explicitly enumerated in the model.

R-dimensions – **trapeziums** – encode concepts that are variables or classes and the graphic structures that allow values of the concept to be depicted. The example in Figure 2 includes an R-dimension for families of airplanes and R-dimensions for variants of specific families. Each family has a three-digit code graphic object and variants have more generic alphanumeric strings. R-dimensions capture the quantity scales for both the concept and graphical object, which are recorded on the right of the R-dimension icons by letters for Nominal (N), Ordinal (O), Interval (I) and Ratio (R) scales. *R-schemes* – **rectangles** – encode structures that are built from other parts of the representation. Figure 2 uses an R-scheme to combine the constant symbol for Airbus, "A", with the alphanumeric codes designating model families, and so captures the entire aircraft model number.

(3) An interpretation of a representation is a network of these schemas, which are linked by two types of connectors: *hierarchy* and *anchor*. Hierarchy connectors associate two schemas with a *reification* (parent–child) relation, where some aspect of the parent concept is inherited by the child; for instance, an R-symbol is a value of its parent R-dimension. In general terms, the hierarchy links allow RISN networks to encode the associations as schemas on the same level in a network and to encode specialization or generalization of concepts as paths down or up the network, respectively. Through its hierarchy connectors, the *Aircraft families* R-scheme sets up a structure for all aircraft that prefixes "A" to the 3-digit number for the family.

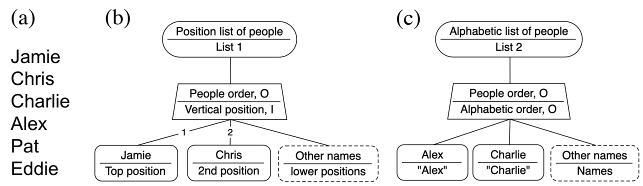

Figure 3. (a) List of people; RISN models of (b) position order list (O-I); (c) alphabetic ordering (O-O).

Anchor connectors are associations between schemas in which the child schema establishes a context for new concepts that are not subsumed by the parent schema. In the example, the R-symbols for the A330 and for the A220 families spawn new R-dimensions for variants that are unique to each of those families.

(4) The last idea is *Idioms* (Stockdill et al., 2022). Idioms are certain substructures of RISN models that are common across different domains and representations. The idioms serve specific representational functions, such as: filtering or specializing concepts through a sequence of R-dimensions; combining concepts in a combinatoric manner by selecting values from multiple R-dimensions.

RISE is a web-based editing tool with rich functionality for building RISN models, including the automatic checking of the syntactic correctness of the network diagrams (Stockdill et al., 2022). All the diagrams of the models in this paper were built in RISE.

Modelling Indexing Systems

To address the second question about index system diversity, RISN models of many indexing systems from each of the four classes were built using the RISE tool following the method of Stockdill et al. (2022). Typical models for each class are presented and the diversity within a class explained in terms of possible variations to the typical model.

Metric indexes

Superficially, metric indexes are simple, consisting of a single property against which an instance is classified or measured (Figure 1a). However, RISN models of metric indexes reveals many potential sources for variations.

To introduce RISN models for metric indexes Figures 3b and 3c show two alternative models for the list of people (instances) in Figure 3a. Both RISN models consist of an R-dimension with associated R-symbols, including a class R-symbols for names not explicitly shown by the models. Figure 3b is an interpretation in terms of position in the list. Figure 3c interprets the names alphabetically, ignoring the positional ordering. As required by RIST, the quantity scale of the concept and graphic objects of the R-dimensions is given. In both cases the concept is a sequence of people, so the scale is ordinal. In Figure 3b the vertical position is considered in this interpretation to be an interval scale, whereas for the alphabetic interpretation the scale is just ordinal.

As a metric index is basically an R-dimension, it is a trivial idiom. More interestingly, what varies across RISN models

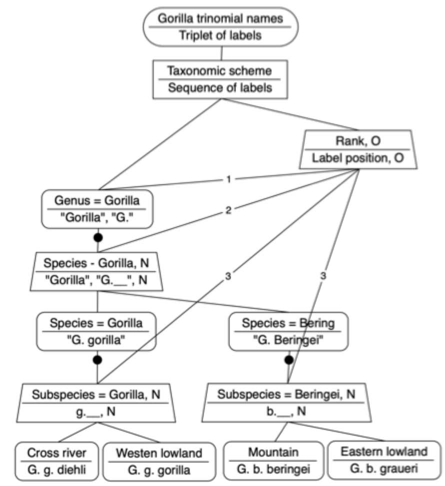

Figure 4. RISN model for gorilla trinomial name taxonomic codes.

of metric indexes are the pairings of quantity scales across the concept and the graphic object. This defines sixteen varieties of metric indexes from all the combinations of assignments of quantity scales to the concept and graphic objects. The examples listed in Table 1a are themselves indexed with pairs of letters, for the quantity scale of the concept and the graphic object, respectively. At one extreme, N-N metric indexes establish identities by the pairing of items on nominal scales (e.g., croquet set). At the other extreme, measuring instruments are often R-R metric indexes: an ammeter registers continuously varying amounts of current on a continuous display that begins from zero. (Together N, O, I, and R constitutes an ordinal metric index for quantity scales.)

The combinatorics of pairs of quantities scales provides part of the explanation of the diversity of this class. Even the simplest design of metric indexes can draw on sixteen possible ways to assign scales to the concept and graphic object, of which examples for eight are given in Table 1a. In addition, the property of metric indexes may be encoded by multiple graphics each with a quantity. The index I-I*2+ in the last entry in Table 1a indicates the case of the composition of multiple interval quantities scales, such as the different sets of letters for unit, 10 and 100 numerals in the Greek numeration system.

Taxonomic codes

Figure 1b shows how instances in taxonomic codes are formed by a succession of properties organized in levels in a tree-like structure, with disjoint subclasses for each class within the tree: the paths down the tree to each instance is unique.

Figure 4 is a RISN model for the interpretation of the zoological trinomial code for gorillas (Mittermeier, Rylands & Wilson, 2013), which exemplifies this class. The overall diagram is the representational system for the whole taxonomy (a tree diagram of the gorilla genus). Each leaf of the diagram is an instance (one trinomial name; e.g., *G. g. gorilla*). The anchoring links from an R-symbol to its child R-dimensions encode the idea that a sub-class of a member of a class is

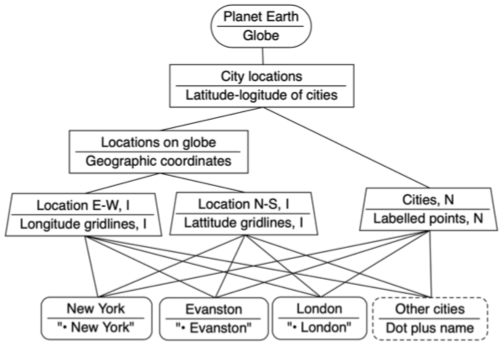

Figure 5. RISN model for the globe triangulational coordinate system applied to locations of cities.

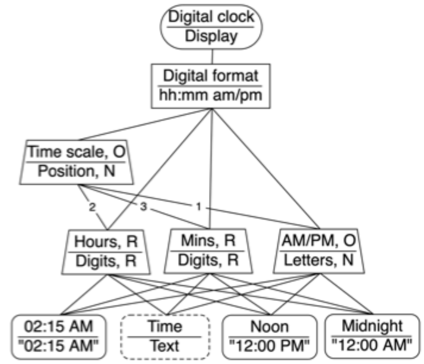

Figure 6. RISN model for a digital clock mereological coordinate system.

unique to that member: the two subspecies of *G. beringei* are unique to it, and not found under *G. gorilla*. The ordering of the labels for the taxonomic ranks is encoded by an R-dimension.

Although the depth of the underlying tree of the trinomial codes is fixed, other taxonomic indexes relax the assumption that the branches in Figure 1b are the same depth. For instance the number of levels in breadcrumb navigation trails or sub-genres may vary from path to path.

Triangulational Coordinate Systems

As the name suggests, the class of index systems identifies an index using values from a selection of properties to identify an instance in the overall space defined by the properties (Figure 1c). In Stockdill et al. (2022) these are *coordinate systems* idioms, which come in two subclasses: *explicit* and *implicit*. In an explicit system one R-dimension specifically identifies a set of instances to interpret as datapoints whose values on the other R-dimensions are to be determined. Figure 5 is an example for cities on the globe located by a geographic coordinate R-scheme that deploys latitude and longitude R-dimensions. The cities R-dimension is the nominal list of instances. In an implicit coordinate system, all the property R-dimensions have the same status; any can be treated as providing instances whose values are to be found in the representation. For example, consider a line graph. Cheng (2020) gives an example of a "monster" graph found in thermodynamics with four R-dimensions that define a coordinate space that includes a subspace with three further R-dimensions. A value on any one of the R-dimensions may be treated as an instance and used to look up values of the other R-dimensions.

Among information visualizations in Table 1c, the graphs and charts are readily interpreted under this idiom, but traffic lights may also be similarly interpreted. They consist of three nominal R-dimensions that each has two values (on/off) and that are distinguished by two metric indexes for colour (nominal) and by position (ordinal). Of the eight possible combinations of illuminations only three are meaningful instances (or four depending on country). Among games, boards with grids are obviously this type of system, but playing cards are triangulational systems too. Each card is an instance and the indexing system is interesting because it is typically interpreted as three coordinating R-dimensions, one for the suits of cards and two that run in parallel, partially: rank (ace to king) and value (e.g., 1 to 11, court cards=10). This richness of indexing is exploited in many well-known card games. This class also includes tools (4th list in Table 1c). A user of a lever-handle mixer faucet/tap may predict or coordinate the desired flow rate and temperature of the water (instance) by separate altitude and rotation movements of the handle. The foot pedals of a manual/stick shift car is a triangulational coordinate system similar to traffic lights: the combined values from three R-dimensions define a variety of operating states (instances). However, unlike traffic lights, all eight combinations of depressed/released values of the pedals are potentially meaningful interpretations, although not to all drivers. For example, the *heel-toe* technique that involves depressing all three pedals simultaneously is an advanced gear change skill taught to racing drivers.

Two major sources of variation in the class of index systems are the number of R-dimensions and differences in the concept and graphic object quantity scales. The number of R-dimensions for each example in Table 1c are in the trailing parentheses. Each R-dimension within a triangulational coordinate system is essentially an embedded metric index, so provides one of 16 different pairings of quantity scales as noted above.

Mereological coordinate systems

This class of index systems is similar to the triangulation coordinate systems as each index draws values from a set of properties. However, mereological coordinate systems differs in that the properties are ordered.

As an example, the RISN model for a digital clock is shown in Figure 6. In the bottom row, the R-symbols are examples of instances. All R-symbols *coordinate* one value drawn from each R-dimension in the layer above; a minute, an hour and an AM/PM value. The levels of the component R-dimensions in the mereological system are encoded by the timescale R-dimension. Although the levels are ordered conceptually, this order is not preserved by the positions of their graphic objects.

Like the triangulation coordinate systems, the major sources of variation are with respect to the number of R-dimensions and the possible pairings of quantity scale to concepts and graphic objects. A form of variation unique to this class is where the ordering of coordinate R-dimensions is reflected in the display. For the clock display it is not: the order is hour-minute-AM/PM. In contrast, the ordering of the magnitude of each digit in Hindu-Arabic numbers is specifically encoded by their individual position: e.g., 100s-10s-units (Zhang & Norman, 1994b).

Hybrid and composite index systems

RISN models for the four classes for indexing systems have been introduced. We do not claim that these indexing system idioms are exhaustive; other specialist systems are likely to exist.

Further, variants of the proposed classes of indexing systems can occur. For instance, the ancient Greek number system (Table 1a) is a hybrid of metric indexes, with a number consisting of letters drawn from sets of letters for units, tens and hundreds that are disjoint, and when a particular order of magnitude is zero the expression omits a letter (e.g., 202= $\sigma\beta$). Changing the order of the letters does not change the value of the represented number. The Airbus naming convention (Figure 2) illustrates other form of hybridity, with the addition of a fixed prefix label on all instances ("A" for Airbus).

Index systems may be composed. The Airbus naming example appends metric indexes to the overall taxonomic code. The N-N metric index for aircraft subtypes is anchored to the A330 R-dimension – compare the bottom left sub-tree in Figure 2 with Figure 3c.

Discussion

Index systems are an important class of cognitive artifacts because they are not only ubiquitous but are also diverse. Given this, two questions were posed: (1) What are the common underlying forms of index systems in general? (2) How should we explain the substantial diversity of specific types of index systems?

In response to the first question, we identify four common classes of index systems: (1) *metric indexes*; (2) *taxonomic codes*; (3) *triangulational coordinate systems*; (4) and *mereological coordinate systems*. Definitions of them were provided and illustrated in Figure 1. RISN models of typical examples of the systems were built, respectively Figures 3, 4, 5, and 6. The models are RIST predictions of the possible schematic memory structures that the external representations of the index systems may invoke in users of the systems. Interestingly, the structure of each model exploits quite distinct RISN idioms, which are basic patterns of RISN network components. For instance, taxonomic codes anchor R-dimensions under R-symbols to encode hierarchical levels. In contrast, in both the triangulational and mereological systems, R-symbols coordinate values from R-dimensions, although they differ in whether the R-dimensions are ordered or not.

To answer the second question, we built RISN models for other examples of index systems in each class (but not presented here in detail) in order to examine how they differ. One major source of diversity for three of the index systems arises from the possible combinations of types of quantity scales. The number of R-dimensions exhibited by the models is another source, which is applicable to the taxonomic, triangulation and mereological systems, and also to hybrid metric indexes.

As with any form of cognitive modelling in cognitive science, there is a possibility that theoretically unconstrained degrees of freedom, which are inevitably present in any modelling approach, may have been unwittingly exploited in the design of the models. We do not consider that it to be a serious risk, for two reasons. First, the RISN models are consistent with the abstract informational structures proposed for each of the classes of index systems and differences among them are all permitted variants under RIST. Second, a good number and range of examples of index systems are included in Table 1, which all neatly conform to the abstract information structures. Nevertheless, empirical studies on the contents of experienced users' interpretations of the indexing will be required to fully evaluate the models.

Diverse indexing systems, from many domains, have been modelled. The functions they support span a broad range of complexity; from the simple pairing of objects (list of names, Figure 3a), through to taxonomic classification (Figure 4), numeration systems, control systems (pedals, taps), and on to complex multidimensional information visualizations. For some systems their categorization was not immediately obvious, and their eventual idiomatic classification initially seemed counter intuitive, but their models revealed interesting parallels to systems for quite different domains. In other words, through the construction of RISN models, common interpretative structures were discovered across representations that are graphically distinct but that might be cognitively equivalent. Further, the set of schemas and relations among schemas was sufficient to produce sensible models of interpretations of all the chosen examples and, importantly, permitted clear definition of the four types of index systems. Thus, in a reflexive fashion, all this appears to provide some further evidence for claims about the validity of RIST and the utility of RISN.

To conclude, we note that the RISN models of the four classes of the index systems provide a foundation for further work that may address the efficacy of different designs of index systems. Zhang & Norman (1994b) and Chrisomalis (2020) have examined the relative qualities of index systems from computational and expressiveness perspectives, respectively. Larkin & Simon (1987) demonstrate the benefits of locational indexing in diagrammatic formats over linear concatenation in sentential formats. As RISN models examine the overall representational structure of index systems – the integration of conceptual and graphical structure – the approach may have some potential to provide general accounts of the efficacy of indexing systems, which are domain and format independent, using general metrics of the structure of RISN networks, such as those proposed by Cheng et al. (2024).

Acknowledgments

This work was supported by EPSRC grants EP/R030642/1, EP/T019603/1, EP/T019034/1 and EP/R030650/1.

References

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4*(2), 75-100.

Cheng, P. C.-H. (1996). Scientific discovery with law encoding diagrams. *Creativity Research Journal, 9*(2&3), 145-162.

Cheng, P. C. H. (2020). A Sketch of a Theory and Modelling Notation for Elucidating the Structure of Representations. In A.-V. Pietarinen, P. Chapman, L. Bosveld-de Smet, V. Giardino, J. Corter, & S. Linker (Eds.), *Diagrammatic Representation and Inference* (pp. 93-109). Cham: Springer International Publishing.

Cheng, P. C.-H., Stockdill, A., Garcia Garcia, G., Raggi, D., & Jamnik, M. (2022). Representational Interpretive Structure: Theory and Notation. In V. Giardino, S. Linker, R. Burns, F. Bellucci, J.-M. Boucheix, & P. Viana (Eds.), *Diagrammatic Representation and Inference* (pp. 54-69). Cham: Springer International Publishing.

Cheng, P. C.-H., Garcia Garcia, G., Raggi, D., & Jamik, M. (2024). A human information processing theory of the interpretation of visualizations: demonstrating Its utility. In *CHI 2024, ACM Conference on Human Factors in Computing Systems*: Association of Computing Machinery.

Chrisomalis, S. (2020). *Reckonings: Numerals, Cognition, and History*. Cambridge, MA: MIT PRess.

Cleveland, W. S. (1985). *The Elements of Graphing Data.* Monterey, CA.: Wadsworth.

Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analysing scientific data. *Science, 229*, 828-833.

Collison, R. L. (1972). Indexes and indexing. London: Ernest Benn Ltd. Collison, R. L. (1972). Indexes and indexing.

Farrow, J. F. (1991). A cognitive process model of document indexing. *Journal of documentation*, 47(2), 149-166.

Giardino, V., Linker, S., Burns, R., Bellucci, F., Boucheix, J.-M., & Viana, P. (Eds.). (2022). *Diagrammatic Representation and Inference: 13th International Conference, Diagrams 2022, Proceedings* (Vol. 13462): Springer Nature.

Glasgow, J. I., Narayanan, H., & Chandrasekaran, C. (Eds.). (1995). Diagrammatic Reasoning. Cambridge, M.A.: AAAI/MIT Press.

Kałużyńska, I. (2015). Traditional Chinese Generation Names. *Onomastica, 59*, 107-121.

Kirsh, D. (2010). Thinking with external representations. *AI & Society, 25*(4), 441-454.

Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology, 3*, 185-226.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.

Kunda, M., McGreggor, K., & Goel, A. K. (2013). A computational model for solving problems from the Raven's Progressive Matrices intelligence test using iconic visual representations. *Cognitive Systems Research, 22-23*, 47-66.

Mittermeier, R. A., Rylands, A. B., & Wilson, D. E. (Eds.). (2013). *Handbook of the Mammals of the World. Volume 3. Primates*. Barcelona: Lynx Edicions.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Catergorization.* (pp. 259-303). Hillsdale, N.J.: Lawrence Erlbaum.

Peebles, D. J., & Cheng, P. C.-H. (2003). Modelling the effect of task and graphical representations on response latencies in a graph-reading task. *Human factors, 45*(1), 28-45.

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), *Artificial Intelligence and the Future of Testing* (pp. 73-126). Hillsdale, NJ: Lawrence Erlbaum.

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Educational Psychology, 91*(690-702).

Shimojima, A. (2015). *Semantic properties of diagrams and their cognitive potentials*. Stanford, CA: CSLI Press.

Stockdill, A., Garcia Garcia, G., Cheng, P. C.-H., Raggi, D., & Jamnik, M. (2022). Cognitive modeling of interpretations of representations. In J. Macbeth & L. Gilpin (Eds.), Proceedings of the 10th Annual Conference on Advances in Cognitive Systems (pp. 1-20). Arlington, VA: Cognitive Systems Foundation.

Tabachneck-Schijf, H. J. M., Leonardo, A. M., & Simon, H. A. (1997). CaMeRa: A computational model of multiple representations. *Cognitive Science, 21*(3), 305-350.

Zhang, J., & Norman, D. A. (1994a). Representations in distributed cognition tasks. *Cognitive Science, 18*(1), 87-122.

Zhang, J., & Norman, D. A. (1994b). A representational analysis of numeration systems. *Cognition, 57*, 271-295.

Zhang, J. (1996). A representational analysis of relational information displays. *International Journal of Human Computer Studies, 45*, 59-74.