

5-28-2012

Extending Classification Principles from Information Modeling to Other Disciplines

Jeffrey Parsons
Memorial University of Newfoundland, jeffreyp@mun.ca

Yair Wand
University of British Columbia, yair.wand@sauder.ubc.ca

Follow this and additional works at: <https://aisel.aisnet.org/jais>

Recommended Citation

Parsons, Jeffrey and Wand, Yair (2012) "Extending Classification Principles from Information Modeling to Other Disciplines," *Journal of the Association for Information Systems*, 14(5), .
DOI: 10.17705/1jais.00332
Available at: <https://aisel.aisnet.org/jais/vol14/iss5/2>

This material is brought to you by the AIS Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in *Journal of the Association for Information Systems* by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Journal of the Association for Information Systems

JAIS 

Special Issue

Extending Classification Principles from Information Modeling to Other Disciplines

Jeffrey Parsons

Memorial University of Newfoundland
jeffreyp@mun.ca

Yair Wand

University of British Columbia
Yair.Wand@sauder.ubc.ca

Abstract

Classifying phenomena is a central aspect of cognition. Similarly, specifying classes of interest is a central aspect of information systems analysis and design. We extend principles originally developed to guide classification in information systems to the general problem of organizing scientific knowledge. Two fundamental cognitive principles underlie the choice of classes. First, classes should encapsulate inferences about the properties of their instances. Second, collections of classes should provide economy of storage and processing. This leads to a view of classes as carriers of domain knowledge in the form of inferences about situations, rather than containers for information. In this paper, we show how this view, originally developed in the IT context, can be extended to other disciplines, notably the natural sciences. We explain how the principles of inference and economy can guide the choice of individual classes and collections of classes. Moreover, we present a generalized classification-based information processing system (CIPS) model. We propose that scientific theories can be represented by class structures as defined in our model and demonstrate how this can be done by applying CIPS to analyze an example from the philosophy of science literature dealing with nuclear physics. The example demonstrates two advantages of the CIPS approach: first, it can provide a simpler, more scalable, and more informative account of the phenomena than a competing approach (dynamic frames); second, the resolution of inconsistencies between theory and observation can be framed in terms of changes to classification structures, and the principles can even guide such changes.

Keywords: *Classification Theory, Conceptual Modeling, Scientific Knowledge, Information Processing, Philosophy of Science.*

* Nicholas Berente was the accepting senior editor. This article was submitted on 24th April 2011 and went through four revisions.

Volume 14, Special Issue, pp. 245-273, May 2013

1. Introduction

Classifying phenomena is an innate and vitally important human cognitive activity; hence, it pervades many fields of human endeavor. In information systems analysis and design, determining classes or types that represent similar phenomena of interest is a common activity¹. In this paper, we extend ideas about classification we developed in the context of conceptual modeling in information system analysis and design to inform classification issues in organizing and reasoning about scientific knowledge.

In scientific disciplines, a classification structure (a set of classes for a domain of phenomena) can constitute both a framework by which theory is developed and a mechanism by which the structure of a domain can be conceptually modeled and understood. Consequently, classification systems play an important role in the philosophy of science (Andersen, Barker, & Chen, 2006; Bowker & Starr, 2000; Ritvo, 1997). Classification of domain phenomena enables scientists to express regularities, and thereby supports hypothesis generation from observations. In our view, the success of these activities is intertwined with the quest for good classification structures.

Guidelines that can assist analysts, designers, and users in choosing appropriate classes can reasonably be expected to improve the quality and usability of information systems (IS). Parsons and Wand (1997, 2008b) propose a theoretical foundation to derive rules to guide class selection in IS applications. In line with the idea that data modeling depends on both ontological and epistemological considerations (Hirschheim, Klein, & Lyytinen, 1995, p. 58), their approach makes explicit assumptions on both aspects. The ontological aspect implies what exists and the epistemological aspect applies cognitive principles to guide how the information about what exists can be organized in classes. The ontological knowledge is operationalized as a domain of phenomena, where knowledge about each phenomenon is specified by a set of properties. The epistemological view is that information (in terms of properties of phenomena) is organized in classes so as to attain cognitive utility based on two principles: inference and economy. The derived rules provide a theory-based approach for identifying classes in conceptual modeling, which is in contrast to prior academic and practitioner research that offered only informal and intuitive guidance in selecting classes to model a domain (Parsons & Wand, 2008b).

In this paper, we further develop insights from the study of classification in conceptual modeling in the information systems field, and examine their implications for other disciplines. In particular, we develop a general information processing model based on classes, and show how it can be applied to scientific knowledge.

We proceed in three steps, and present principles and examples for each step. First, in Section 2, based on fundamental work on the role of classification in cognition, we focus on what a class is. In Section 3, we apply the notion of class to understand a classification controversy in astronomy (extending Parsons & Wand, 2008a). In Section 4, we consider additional principles that can guide the selection of collections of classes to conceptualize a domain. In Section 5, we illustrate these principles using examples from astronomy and citizen science. Building on principles of classification, we show how collections of classes can be used to process information. We propose a general classification-based information processing model (CIPS) in Section 6 and demonstrate its application in Section 7. In Section 8, we explain how this model can be used to represent scientific knowledge and demonstrate this use via an example taken from nuclear physics that was previously analyzed using another approach (Andersen et al., 2006). Our analysis shows that the CIPS-based representation provides both a simple account of changes of scientific knowledge in terms of

¹ For example, in conceptual modeling, classes represent relevant concepts from the application domain. In database design, classes (e.g., relations) are the primary mechanism for structuring data. In object-oriented design, classes are templates encapsulating structure (data) and behavior (methods) of software units. In Semantic Web applications, formal ontologies provide standardized terminology and support information sharing by specifying the "classes of objects and relations that exist in some domain" (Chandrasekaran, Josephson, & Benjamins, 1999). The structure of menus in a software application represents a classification of commands organized to facilitate user actions. Finally, folder structure is used to organize files and can have a direct impact on users' ability to quickly locate them later.

classification structure and a mechanism to generate questions that can be asked to resolve discordances between observations and prior thinking.

In short, this paper shows how research originally undertaken to support information systems analysis and design can contribute to scientific understanding and to the notion of theory in other fields of research.

2. Cognitive Principles Underlying Classification

Classification is a pervasive human activity. It has been a longstanding research focus in the cognitive sciences, where the classification (or categorization) of phenomena has been recognized as critical to human survival (Lakoff, 1987). We use the general term “phenomenon” to indicate anything that can be identified and described and, hence, that might be classified. While usually rooted in observations and experiences, phenomena reflect human perceptions about material things, imagined things, states, events (or sequences thereof), laws or rules, symbols, and scripts composed of symbols.

Cognitive research on classification has proceeded largely on the basis that the critical purposes of classification are to make efficient (economical) use of cognitive resources and to enable inferences (Rehder & Burnett, 2005; Rosch, 1978; Smith, 1988; Yamauchi & Markman, 2000). As classification is a purely cognitive phenomenon, efficiency and inference can be considered the only purposes of classification. Cognitive efficiency refers to the goal of maintaining the minimal necessary information about a phenomenon for possible future use; inference refers to the ability to deduce information about a phenomenon as a result of identifying it as an instance of a class (Rosch, 1978). Together, efficiency and inference enable humans to manage and use the large amount of information to which they are exposed. They provide the guiding criteria underlying the process of abstracting from individual phenomena in a domain to distinguish concepts that are useful and meaningful from those that are not.

2.1. Cognitive Efficiency²

Consider a domain of interest comprised of a set of phenomena about which we want to maintain information. A specific phenomenon will be called an “instance”. Each instance can be described by a set of properties. We use “property” to refer to any observable (directly or indirectly, or even imagined to be observable) characteristic of an instance. A property can be conceptualized as a statement that can be made about an instance at a given time. We term such a statement, which might reflect observation, knowledge already possessed, communicated information, or inferences based on any combination of these, an “attribute”. While humans communicate via attributes (e.g., “this car is red”), in the remainder of this paper, we assume that all such statements reflect properties of interest, and will simply refer to “properties”. In essence, the choice of properties reflects the user’s point of view.

Efficiency can be related to resources used for either storage or processing. We first consider efficiency in terms of economy of storage. Suppose a number of instances possess a common “bundle” of properties. Rather than storing all properties for each instance, it might be more economical to store a description of the bundle (in terms of properties) once and, for each instance possessing it, store a reference to the bundle. Economy of storage can also lead to economy of resources needed to transfer information (economy of communication) because one can refer to the bundle of properties by naming it, rather than by referring to each property in the bundle (provided the recipient of communication understands the bundle in terms of the same properties).

In addition to storage economy, it might be faster to retrieve a bundle from memory than to retrieve the properties about each phenomenon individually. Moreover, if several processing components need to share information (either in artificial or in biological systems), it will require less resources to move information organized into such bundles. We suggest that such a (potentially economical to store, communicate, process, and move) bundle defines a “category”, and the related phenomena are

² The definitions in this section are adapted from Parsons and Wand (2008b).

the instances of the category. This view of categories reflects a “classical” perspective, in which each instance of a category possesses all properties defining the category³.

While cognitive economy can be achieved by creating categories, whether it is attained or not and to what degree depends on the size, prevalence, and frequency of use of the bundle of properties used to define a category. Note that a given phenomenon can be an instance of several categories. In addition, some categories may subsume others, meaning that categories can be arranged in a lattice (where multiple inheritance is possible) reflecting specialization and generalization relationships.

As indicated earlier in this section, in addition to economies related to storage and retrieval, a category can provide economy of communication if the involved parties have a shared understanding of that category. Categories can also trigger additional knowledge acquisition. Knowing that a phenomenon belongs to a category might suggest questions that can be asked about it. For example, if a certain insect has wings, one can ask “How many wings?”.

The last example illustrates a common situation in which instances have “specific values” (e.g., number of wings) related to a “generic property” (e.g., having wings) that is part of the category definition. The notion of generic property is quite common because we often abstract properties from specific values to more general ones. For example, because every person has an individual weight, we think about persons as ‘having weight’⁴. Similarly, specific objects reflecting a given wavelength are assigned the attribute ‘red’. We can generalize to a property of ‘being red’ and even to ‘having a color’. Less obvious examples include: ‘having flown a certain mileage’ and ‘having flown a certain number of flights’, which can be generalized to ‘being a frequent flyer’; and ‘high value of purchase’ and ‘being a member of a loyalty club’, which can be generalized to ‘preferred customer’. In these cases, properties that appear to be different are generalized to the same more general (abstract) property. We call the former “manifestations” of the later (Parsons & Wand, 2003). Thus, property abstraction is formalized in terms of property manifestation. Property abstraction creates hierarchies of properties, whereby an abstract property is manifested by several more specific properties, and even lattices, whereby a property can be abstracted to more than one “immediate” higher level property. Importantly, classification is often based on the more abstract properties while instances possess the manifestations.

However, the definition of a category might also include specific values of more general properties. For example, a common social category is “senior citizen”, which is possibly defined by a specific range of values (at least a certain age) of the property ‘having age’. In addition, it is not necessary that values be exact; they may instead be qualitative (e.g. ‘excellent student’, ‘large dog’, or ‘patient at risk’). We note that individual instances, in addition to having more specific values of their category-level properties, usually also possess additional properties (indeed, there are ontological positions that claim this must be the case). This, in particular, can give rise to subcategories, when several instances of a category share properties in addition to those defining the category.

For a given domain of interest, the number of possible categories can be very large. Maintaining large collections of categories can entail extra maintenance and search costs that might offset the economy afforded by individual categories. We claim that a second principle – “inference” – is the criterion that determines which categories can be useful and will be more likely to be maintained over time.

2.2. Inference

Assume that membership in a category can be determined by a subset of the properties defining it. For example, instances possessing the properties ‘live only in sea’, ‘have fixed body temperature’, ‘breathe air’, and ‘have blubber’, can be identified as instances of the category “sea mammals” by the

³ This view can be extended to accommodate other approaches to defining classes – in particular, the prototype and exemplar models (Rosch, 1978; Smith, 1988). Other extensions are possible. For example, properties associated with a category can be assigned probabilities.

⁴ Throughout the article, single quotation marks indicate property (attribute) names and specific values. For example: ‘having color’, ‘red’ or ‘being red’. Category names will be indicated by double quotation marks (e.g., “senior citizen”).

first two properties. In such cases, one can infer that an instance possesses some properties by knowing that it possesses some others⁵. Such inferences are useful because they enable one to know that some instances possess certain properties without having to expend the effort or incur cost or risk to observe these properties directly.

We term such a category, defined in terms of a set of properties and inferences, a “class”, and the set of phenomena possessing these properties the instances of the class. A class is useful in that it provides more knowledge about an instance than is needed to determine that it is a member of the class. For example, “if this is a lion, then it is dangerous” or “if a person is a preferred customer, then she is eligible for a discount and free delivery”. We term a subset of properties sufficient to determine class membership a “base”. For example, high mileage or many flights travelled can each be a base for the class “frequent flyer”, for whose instances we can infer eligibility for various benefits. The existence of a base and the existence of inferences that are common to all instances of a category are equivalent (Parsons & Wand, 2008b).

To summarize, we use cognitive inference to suggest that what makes a class useful is that additional information (properties) can be deduced once an instance is identified as a member of the class. The usefulness of a class is in implying additional properties.

There is an important distinction between categories and classes. A category simply reflects a repeating pattern of properties possessed by some phenomena. In contrast, a class also implies predictable interrelationships among the properties of its instances. In that sense, classes capture regularities that might reflect either natural or artificial laws.

Classes provide benefits additional to those of categories. It is sufficient to acquire knowledge of only some properties of a phenomenon to identify it as an instance of a class and thereby infer the remaining class properties. This will be beneficial if acquiring the inferred properties in other ways (e.g., observation or communication) is more difficult, time-consuming, dangerous, or costly than inferring them. Making inferences is akin to processing (in cognition, other biological systems, or artificial systems). Hence, inferences encapsulate the idea that in some situations processing has advantages (is more efficient or effective) over observation or communication in obtaining information. Table 1 provides a summary comparison of categories and classes. Recall the above distinction between general properties and specific values or manifestations. Often, instances of a category possess specific values related to more generic properties used to define the category. One type of inference of particular interest is that which links the values of some properties to the values of other properties. Such links can exist even if the specific values are not exact. For example, the normal life span of a dog depends on whether it is ‘large’ or ‘small’.

Table 1. Comparing Categories and Classes

	Definition	Identifying instances	Usefulness
Category	A set of phenomena all possessing a common set of properties.	A phenomenon is an instance of a category if it possesses all properties defining the category.	Storage efficiency. Communication efficiency.
Class	A category where some properties can be inferred from other properties.	A phenomenon is identified as an instance of a class if it possesses some of the class properties (the base).	Ability to infer more information about a phenomenon once it is identified as an instance of the class. Can reduce effort, time, and risk to obtain information.

⁵ Formally, if “instance x possesses a set of properties Q” implies “x also possess the set of properties P” (where P and Q are disjoint), we say that P precedes Q (or P is inferred from Q). Precedence always exists between specific properties and their related general properties.

For a class to be reliable (that is, its inferences can be believed valid), some “stability” of inferences must exist. In other words, the inferences must be true over a period of time. Otherwise, inferences implied by the class cannot be “trusted”. The duration that determines reliability depends on the nature of the phenomena and on the purpose of classification. In astronomy, the “demotion” of Pluto from the class “planet” in 2006 is a case in which more information became available (about the existence of other celestial bodies very similar to Pluto) without any changes in the world (Parsons & Wand, 2008a). Thus, the time period is related to accumulation of knowledge and to purpose, rather than to any natural process. In contrast, biological classification is subject to both changing knowledge and changing properties over time as a result of evolution (see Parsons & Wand, 2008a, for discussion).

The importance of inference in classification can be understood in terms of the “survival value” it affords. Survival value arises from the ability to combine observations efficiently with preexisting knowledge to make inferences about future good or bad situations. Such situations can arise as a result of actions by the environment (nature), by others, or by us. Inferences about future situations, namely predictions, can enable us to determine the impact of taking (or avoiding) action on enhancing our resources (e.g., acquiring food, warmth, shelter) or keeping us from danger (e.g., avoiding being eaten). In particular, inference and prediction might save critical time in situations of immediate danger. This does not mean all inferences we make support survival, but rather that the cognitive ability to infer likely arose due to its survival value. Thus, principles underlying classification can be derived based on the role of classification in survival.

Storing and searching prior knowledge and making predictions also require resources (memory, time, and effort). We therefore posit:

The choice of classes is driven by the need to balance the usefulness of inferences (effectiveness) against the resources required to store and process information to support such inferences (efficiency).

Based on this view of classification, it makes sense to expend the resources required to form a category only if, as a result, we can:

- (1) Infer information about the phenomena belonging to the category beyond that used in identifying those phenomena as members of the category (namely, the category is a class), and
- (2) Reduce the resources required to obtain, maintain, communicate, and process information about these phenomena.

These are the two main principles we will use to develop the ideas of what constitutes a good classification, and how to use classes for information processing. In essence, these principles argue that classes may be added to a classification structure only if they provide some additional benefit in terms of inference or effort required to classify, and if they do not include only redundant information. These principles can be considered an “Occam’s razor” in classification systems, which reflects usefulness and economy.

We distinguish two types of inference (Holland, Holyoak, Nisbett, & Thagard, 1986). First, it is beneficial to spend the minimal effort needed to acquire all relevant aspects of a given situation. This indicates the value of making inferences from some known properties to other properties at the same time (“synchronic” inferences). Second, once sufficient information about a situation is known, inferences can be made about possible situations at a different time (“diachronic” inferences). The two types of inference are complementary in that both might be needed to contribute to survival (or, more generally, to the usefulness of classes). First, given some properties of a phenomenon, more properties that hold for the same time might be inferred without having to expend the effort or risk required to acquire them. Second, given the full set of properties, diachronic inferences can be made

with respect to some properties at a different time (usually in the future). Finally, synchronic inferences may be again used to deduce additional properties for the different time.

In Section 3, we illustrate the concepts of category and class and their role in scientific understanding by reviewing an example from a recent scientific controversy: the definition of planet adopted by the International Astronomical Union in 2006 (Parsons & Wand, 2008a). We chose this example because it comes from a completely different domain than information systems and because it received much popular attention.

3. Category Versus Class in Science

3.1. The Status of Pluto

In 2006, the International Astronomical Union (IAU) voted to define planets in our solar system as follows: “A ‘planet’ is a celestial body that: (a) is in orbit around the Sun, (b) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, and (c) has cleared the neighborhood around its orbit” (IAU, 2006). As a result of this vote, Pluto was no longer considered a planet in our solar system, but was re-categorized as a “dwarf planet”.

If the above three statements are all the common properties of the (newly defined) planets and none is implied by the others, then “planet” is a category. If, however, some consistent relationships exist among some properties of planets, then “planet” is a class as defined above. Consider that all celestial bodies possessing the above properties have some additional properties. For example, their orbits are all coplanar (on the ecliptic) and do not intersect any of the others’ orbits. Thus, an observed relationship is that the three planet-defining properties imply the last two properties. This would not be the case if Pluto was considered a planet since it does not lie on the ecliptic and its orbit intersects that of Neptune. Hence, it seems deep reasons (perhaps related to the formation of our Solar system) might exist to exclude Pluto⁶. In our terms, the (newly defined) concept of “planet” is more than a category. Rather, the new definition constitutes a class in that certain inferences hold from the defining properties to others (coplanar and non-intersecting orbits). Note that this does not mean that Pluto (and other bodies in the solar system) cannot belong to other classes with the planets that share a different common set of properties and inferences. For example, observe that all bodies that orbit the sun and are nearly round have much less eccentricity of orbit than some bodies (e.g., comets) that orbit the sun but are not nearly round. Thus, the controversy sparked by the redefinition of “planet” in such a way as to exclude Pluto is unwarranted. In essence, what matters is the purpose of forming the specific classification. To summarize, whether a certain set of instances will be considered a class depends on whether inferences of interest exist among their common properties.

The Pluto example shows how our rules of classification provide insights about the nature of classification in science. In particular, classification is about understanding the rationale for forming a specific set of classes (identifying the inferences of interest, which reflect the purpose of forming the classes). We now explore principles that apply to collections of classes.

4. Principles For Choosing Collections of Classes

Cognitive efficiency and inference have been used as guiding principles for identifying classes that capture domain knowledge effectively and efficiently (Parsons, 1996; Parsons & Wand, 1997; Parsons & Wand, 2008b). The principles provide two main insights: (1) inference determines what makes a category “useful” (hence a class), and (2) when forming a collection of classes, efficiency helps determine which classes to include.

⁶ Whether such reasons do in fact exist, or can be identified, is a question for astronomers. Our point here is that applying the distinction between category and class helps identify questions (in this case, “are there deeper reasons to exclude Pluto?”) that can be used to guide further research or investigation.

We now describe how inference and efficiency are used to derive specific guidelines for choosing collections of classes in conceptual modeling (Parsons & Wand, 2008b). In Section 2, we define the concepts of category and class. Here, we use cognitive efficiency to suggest principles for forming “good” collections of classes. We then derive specific necessary conditions for a set of classes to adhere to the principles.

Usually, collections of classes are used to represent information about phenomena in a domain. A collection of classes constitutes a model of a conceptual structure (or ontology) of useful abstractions in a domain. The classes in such a structure may be related in a hierarchy or lattice of “subclass” / “superclass” relationships between classes. More than one such set of classes might exist for a given domain of phenomena (Parsons, 1996; Parsons & Wand, 1997).

We now turn to defining “good” collections of classes. The notion of a class was driven by inferences. In contrast, we propose that the choice of a collection of classes should be driven by economy of resources used to identify classes, maintain class definitions, and do the processing required to identify class membership and perform inferences. In this respect, it does not matter whether the processing resources are cognitive or technological. The guiding principle is to minimize resources without loss of relevant information (in terms of instance properties and inferences).

Considering that inferences are a necessary aspect of classification, we posit that a collection of classes chosen should satisfy two types of principles. The first type is completeness—intended to assure that no information is lost—and it has two aspects. First, the choice of classes must assure that for each instance, all relevant properties can be found. This will be done by identifying all classes to which an instance belongs and all specific properties (values) that manifest the class-level properties for these classes. Second, when properties can be inferred from others, for every applicable inference there should be a class for which the inference holds. The second type of principle is intended to assure efficiency and it also has two aspects. The first is non-redundancy—each class must provide information not provided by any other classes in the collection. This information can be in the form of properties and/or inferences. The second is simplicity—when alternatives exist for representing information, the simplest class definitions should be chosen.

For a collection of classes that satisfies completeness, efficiency (i.e., minimizing storage and processing resources) can be assured by following two rules. First, “maximal abstraction” states that whenever two or more classes can be generalized, the generalized class should be included. This means that if two classes (or more) share some (but not all) of their properties and an inference related to these properties, a higher class should be formed with the shared properties. This will be a class because an inference exists among its properties. The Linnean taxonomy (at levels higher than species) embeds this principle. Second, “non-redundancy” states that no class should be included unless it adds value in terms of information (properties or inferences) not provided by other classes⁷. To illustrate, consider two classes: “mammals” and “animals that live only in the sea”. It can make sense to define a subclass of these two “sea mammal” because we can identify an organism as a member of this class without having to observe all the properties needed to identify it both as a mammal and as a sea dweller (specifically, if it has a fixed body temperature and lives only in the sea, it is a sea mammal). In contrast, this will not be the case for “students” who are “surfers” if there are no additional properties of this group beyond being students and surfers. In this case, students who are surfers will be a category, not a class. We term a collection of classes satisfying maximal abstraction and non-redundancy a “class structure”.

The above analysis implies that, when following maximal abstraction, each class needs only to be evaluated with respect to its superclasses as to whether or not it adds new information (properties and inferences). Specifically, it must have at least one additional inference with respect to its superclasses.

⁷ The principle of non-redundancy also implies that, for a class to have properties not included in any other class, it must have an inference not included in any other class (Parsons & Wand, 2008b). If such an inference does not exist, then the effort to identify an instance as a member of the class is exactly that required to observe all class-specific properties and the class provides no cognitive benefit.

Table 2 summarizes the principles and rules for defining class structures. The rules imply that, beginning with any collection, the classes can be examined for possible abstractions (commonalities of properties). Abstractions should be taken to the highest level (minimal number of defining properties), but should preserve the inferences related to those properties. Higher level classes should be formed only if they also contain inferences. As a result of such abstractions, some classes might become subclasses of higher level classes. To eliminate redundancy, each class should have properties in addition to all its superclasses and at least one additional inference relating to these (and perhaps other) properties. Following this procedure will assure that the collection of classes is a class structure.

Table 2. Principles for Defining Class Structures		
Principle	Description	Rule
Completeness	All relevant information about each phenomenon (instance) in a domain should be included	
Properties	Every relevant property of every instance should be included in at least one class	Each class should have at least one additional property with respect to all its superclasses*
Inferences	Every inference about properties of instances should be included in at least one class	Each class should have at least one additional inference with respect to all its superclasses*
Efficiency	Minimize resources used in maintaining and processing information	
Non-redundancy	A class should be included only if it provides information not included in any other class	New information – inferences: A class must have an inference not included in any other class* New information – properties: A class must have properties in addition to all its superclasses*
Minimize resources	Minimize resources required to identify class membership and make inferences	Maximal abstraction: Every inference is included at the highest level of classification (“simplest” class) possible
* These two conditions are related and imply that if a class inherits all its properties and inferences (from its superclasses), it should not be included.		

Finally, given a class structure, guidelines can be established with respect to determining whether or not to add categories to the structure (Parsons & Wand, 2008b). First, the category must be a class (have inferences). Second, all possible superclasses of this class should be identified. Third, the new subclass should be examined to see if it possesses at least one new inference with respect to its superclasses. Prior research proposed specific rules with respect to subclasses in general and to subclasses that are formed by one of two common specialization mechanisms (Parsons & Wand, 2008b). First, classes are often formed as subclasses of several existing classes (e.g., a Teaching Assistant is a Student and an Employee). Second, classes are often formed by restricting the value of a class level property (e.g., a Senior Citizen is a resident who is of age 65 or more).

In Section 5, we consider the implications of these principles for collections of classes that represent knowledge in a scientific discipline.

5. Implications for Class Collections in Science

The proposed principles provide clear guidance for developing class structures. Multiple class structures can be constructed to model a domain of phenomena. Different structures may be useful

for different purposes, and there is no inherently “correct” way of classifying the phenomena in a domain (Parsons & Wand, 1997, 2000). In this section, we examine how these implications help understand the role alternative class structures can play in modeling specific scientific domains.

In the earlier discussion of planets, we consider the requirement that, for “planet” to be a class, it must be possible to identify a body in our solar system as a planet based on a set of properties and infer that the body must possess additional properties by virtue of being a planet. We can generalize this approach to subclassify the planets. Consider the planets (according to the IAU definition) beyond Mars. All share the additional properties ‘large’ and ‘gaseous’. Consider the three planet-defining properties together with the three properties ‘beyond Mars’, ‘large’ and ‘gaseous’. Because Pluto is not a planet, from the three properties of “planet” defined by the IAU, with any one of the other three, we can infer the remaining two. This enables the definition of a subclass of planet: “gas giant”. The same approach can be used to conceive of different ways of classifying the bodies in our solar system. From an astronomical perspective, these classes include planets, asteroids, dwarf planets, moons, and comets. This classification scheme is based primarily on properties related to orbit and shape. However, other valid classification schemes are possible. For example, bodies can be classified by whether or not they have water on the surface (important in inferring their ability to sustain life), their size (important in inferring gravitational characteristics), or their composition (important in inferring whether there are potentially exploitable minerals).

A second example illustrates how a prevailing view of classification in science can impose constraints on the collection and use of data in research. Citizen science refers to the participation of amateur volunteers in scientific research (Silvertown, 2009). In a natural history context, citizen science relies on voluntary contributions of observations of flora and fauna in a region (e.g., Dickinson, Zuckerberg, & Bonter, 2010; Sullivan et al., 2009; Wiersma, 2010). The general focus of attention in such projects is identifying the species of any given observation (Parsons, Lukyanenko, & Wiersma, 2011). This can create a barrier to participation by novices and an impediment to data quality in citizen science projects (Lukyanenko, Parsons, & Wiersma, 2011). In addition, this focus restricts the kind of information collected in such projects and its potential uses. For example, collecting information at the species level is useful for making inferences about the range, distribution, and frequency of biological species in a geographic region. In contrast, if one is interested in examining the impact of an oil spill on birds and fish, classification by species may be far less important than classifying observations based on the condition of specimens and the time and location of observations. Such classifications would enable scientists to make inferences about the degree and consequences of pollution caused by a spill. In addition, a fixed classification structure prevents the collection of data about instances and their properties that might lead to identifying new species (new classes) because fixed classification constrains both the properties that can be collected for observations and the possible combinations of these properties.

Collections of classes are formed with the anticipation that they will be used. Hence, a complete account of classification principles must include, beyond guidance for the formation of class structures, a description of how collections of classes can be “called on” to do useful information processing work. Therefore, in Section 6 we suggest how the principles of economy and inference in classification can provide a foundation for a general information processing model.

6. An Information Processing Model Based on Classes

We propose a generic model for using classes in the context of information processing. Assume a given set of classes has been identified. Information about a phenomenon (in terms of properties) can be obtained in three main ways. First, information can be acquired via direct observation or by communication from external sources (input)⁸. Second, information can be retrieved from prior knowledge (memory). Third, additional information can be inferred from the information obtained by input and retrieval based on identifying the phenomenon as an instance of some classes (processing).

⁸ Communication means any form by which information can be transmitted to an individual from others, such as oral, written, gestures, displayed, and mechanical means. The important feature is that the observation is not done by the classifying agent (person, or natural or artificial system).

Class identification involves comparing known properties of a phenomenon to patterns of properties defining known classes⁹. Acquisition, retrieval, and inference need not always occur, or need not be done by the same agent (individual or system) and can involve several communicating agents. In particular, classifying a phenomenon can be performed by one agent and communicated to another. Additionally, prior classification of a phenomenon can be stored in memory. Thus, communicated and retrieved information about a phenomenon can include both properties and identified categories or classes. Finally, once class assignment has been done, a way to perform inferences must exist. We therefore argue that the use of class structures should be analyzed in a framework allowing for:

- Acquiring knowledge of properties or pre-classification of phenomena by observation or communication
- Storing known properties of phenomena
- Retrieving stored knowledge (properties or prior classification) of phenomena
- Storing known patterns that represent classes
- Storing inferences that can be made about the instances of classes
- Matching known properties to pre-defined patterns (classes), and
- Performing inferences about (additional) properties based on identified classes.

We term a system that provides the above capabilities a ***class-based information processing system*** (CIPS). Figure 1 illustrates CIPS using the common model of input-processing-output. For simplicity, recall is not shown as a process. All solid arrows signify information flow. The dashed arrow refers to the possibility that information stored about a phenomenon already includes identification of some categories (or classes) of which the phenomenon is an instance.

The CIPS model is neither domain specific nor dependent on the mechanisms used to observe, communicate, store, retrieve, and perform inferences. We illustrate this view by considering medical diagnosis in terms of classification. The classes are various medical situations (diseases). Assume a physician sees a patient and observes certain symptoms (properties obtained by observation). The physician also has prior knowledge of the patient (recalled properties and possibly classification into risk groups) and obtains results of some tests (properties obtained via communication). Together, this information is used to diagnose the patient's situation (identify classes). Based on the diagnosis, inferences can be drawn as to prognosis and possible effects of treatment. These involve both synchronous (about current state) and diachronous (about possible future state and impact of intervention) inferences.

Classes can be used not just to infer additional properties. This will be the case when the process of matching known properties of a phenomenon to class definitions results in partial matches. Two possibilities arise. First, a partial match can indicate that more information might be needed to achieve a successful classification. Thus, attempts at classification can trigger recognition of a need for further information acquisition in the form of questions to ask, or observations to make. Medical diagnosis is a good example. A physician might conclude that several possible conditions can match the observations about a patient and order more tests to identify the actual condition. Second, once a phenomenon is recognized as an instance of a category, more questions might arise about specific properties. For example, in citizen science, identifying an organism as a bird can trigger questions about color and length of beak.

⁹ This view can accommodate not just properties as necessary and sufficient conditions for class memberships, but also other approaches such as prototype and exemplar views of classification.

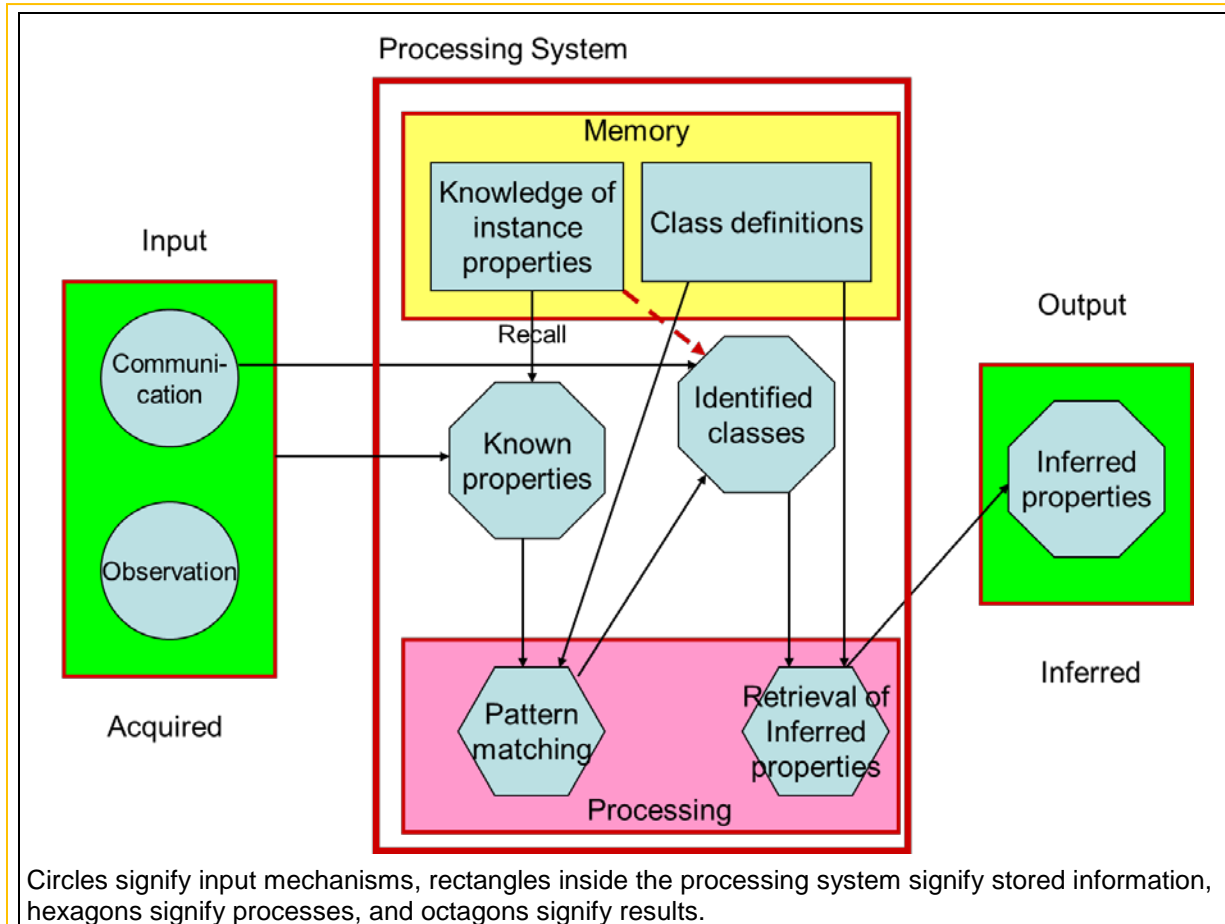


Figure 1. An Input-Processing-Output Model of Classification-Related Inference

Finally, we comment on the use of CIPS as a model for acquiring new knowledge. Consider an observed instance that has a minimum set of properties to be an instance of a class in the classification structure. Assume some other observed properties do not match the properties inferred for this class. The question then arises why this might be. If we rule out errors of observation, the only possible explanation is that, for some classes in the class structure, the inferences are not universal. In other words, the knowledge reflected by the class structure is (at least partially) incorrect and needs to be modified to reflect the new observation. Correcting class structures to match new observations can be considered a feedback loop, where the adjustments are made to the class structure, not to the inputs (which reflect observations assumed to be true). In Section 7, we discuss what modifications can be made to a class structure and use this idea in Section 8 to show how scientific theories can evolve to include new observations.

To demonstrate how CIPS can be used in different domains, we include in the Appendix a brief description of the application of CIPS to the immune system and to a brain structure. In Section 7, we describe an alternative analysis framework from the literature (frames), and compare it to CIPS using a simple example that demonstrates the advantages of CIPS over frames for understanding conceptual models of a domain.

7. Applying CIPS and Comparing it to Frames

Andersen et al. (2006) use the idea of “frames” to describe scientific knowledge. Frames convey views of concepts in terms of a hierarchy of nodes. One level of nodes describes attributes of phenomena that are instances of the concept, and the other describes values of these attributes. Specialized categories are then formed by choosing sets of attribute values. It is assumed that each phenomenon of interest appears in exactly one category (combination of attribute values). Frames allow for substantial flexibility as any aspect of a frame (concept, attribute, and attribute value) can be further elaborated upon as the main concept of another frame. Additionally, frames link concepts, attributes, and values graphically in a clear way. In Figure 2, we present a frame for “vehicles”, developed to illustrate the differences between frames and the CIPS model.

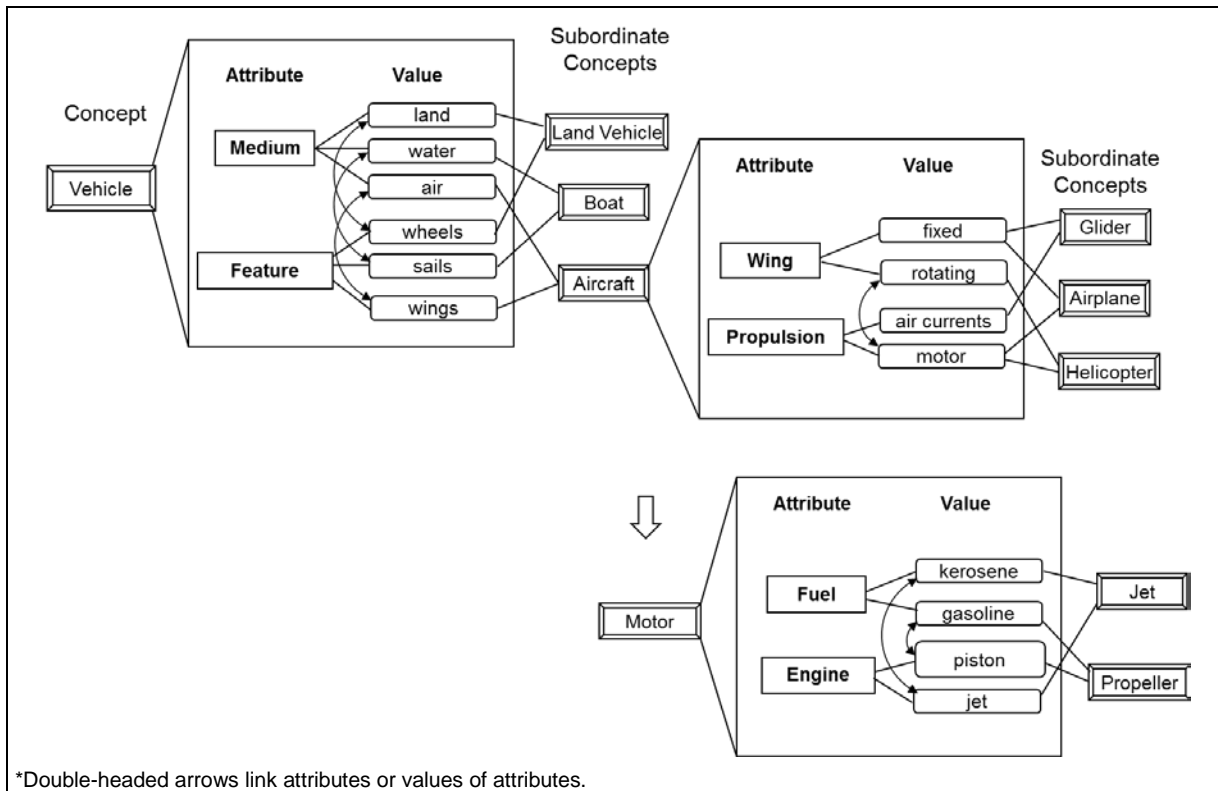
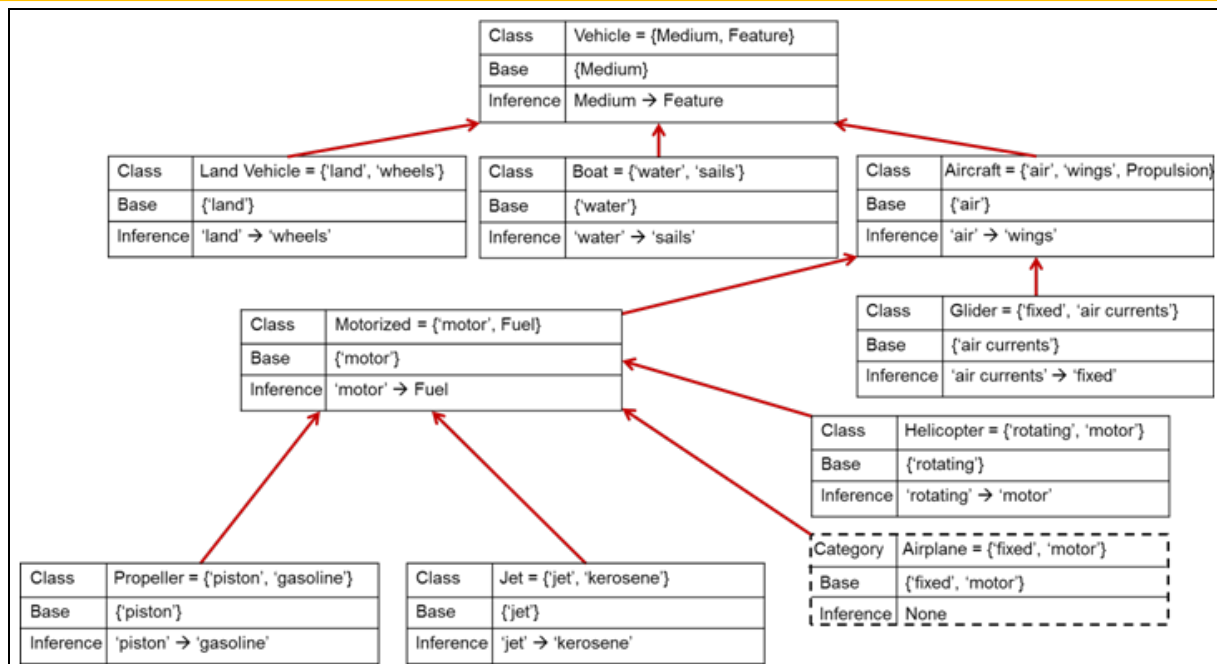


Figure 2. Example of the Use of Frames

Figure 3 shows a CIPS view of the same set of phenomena. There are several important differences. First, frames can refer to any aspect of phenomena (categories of phenomena, attributes of the phenomena, or values of attributes). In contrast, in CIPS categories have a uniform meaning—instances of any category represent the same type of phenomena in the domain. Therefore, categories can be related by subclassification relationships. This is not necessarily the case for frames, as demonstrated by Figure 2. Second, all categories in a CIPS representation are classes—they must have inferences. Note that Figure 3 explicitly specifies the inferences associated with each class and depicts subclassification explicitly and unambiguously.



The following relationships exist among attributes and attribute values:
 Medium: {'land', 'water', 'air'}; Feature: {'wheels', 'sails', 'wings'}; 'wings': {'fixed', 'rotating'}; Propulsion: {'motor', 'air currents'};
 'motor': {'piston', 'jet'}; Fuel: {'gasoline', 'kerosene'}

Figure 3. Vehicle Domain Expressed Using CIPS Concepts

According to Figure 3, there is no specific inference related to instances of “airplane”. Therefore, it is a category, but not a class. The CIPS model requires that we either identify an inference or delete the category. In Figure 4, we have identified one such inference (“runway”—meaning “requires a runway”). This demonstrates a key feature of the CIPS framework—every class requires a new inference.

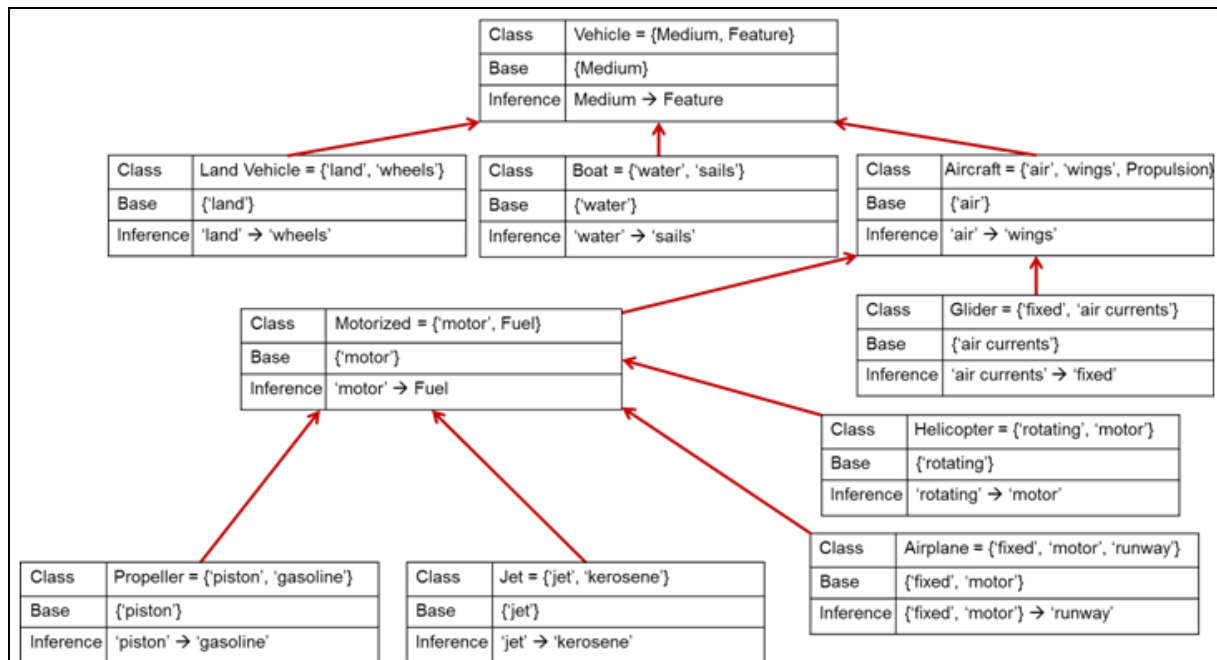


Figure 4. Vehicle Domain Corrected for Inferences

Consider the class structure (with inferences) as a “theory” of the “vehicle” domain of phenomena. We now analyze what can happen if a new instance of vehicle is found that does not fit in any of the existing classes. For simplicity, we focus only on the subtree of aircraft and its subclasses (Figure 5). Assume the following observation has been made: “an aircraft with fixed wings that can land or takeoff on water”. The observed properties are: ‘motor’, ‘fixed wings’, and ‘lands on water’ (no runway).

The CIPS framework provides us with a way to reason about such changes. First, the question arises—since the categories reflect real (observed) phenomena, is it useful to maintain them? The first possibility is simply to eliminate the category “airplane” from the original class structure. In the case of scientific theories, this means that what was considered a natural law is found not to be valid. However, a second question arises—if we do not retain the category “airplane” (or its sub-categories), where would properties such as ‘runway’ be accommodated in a CIPS description of the “aircraft” domain?

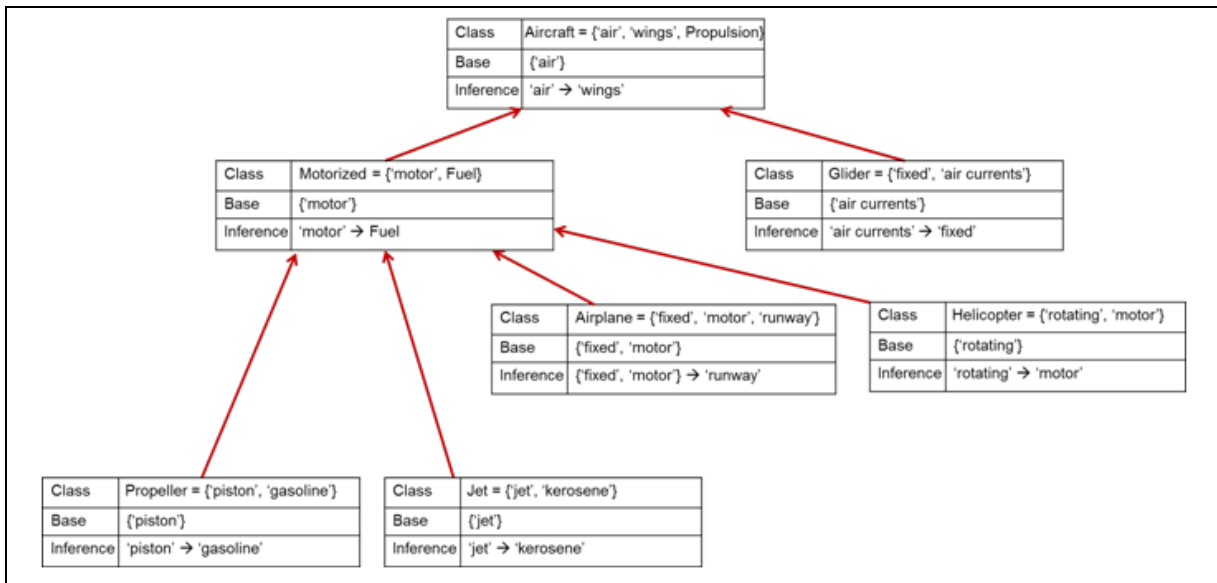


Figure 5. Class Structure of Aircraft and its Subclasses

The critical point is that, for ‘runway’ to be included, it must appear in an inference. Two cases exist: first, it can be used (perhaps together with other properties) to infer another property; second, it can be an inferred property. Consider the first case: ‘runway’ is necessary to know the possible routes of an aircraft (it requires runways to depart and arrive). If we add the property ‘routes begin and end in airports’ (briefly: ‘routes include runways’), then the category (‘motorized’, ‘fixed wing’, ‘runway’, ‘routes include runways’) with the inference {‘motorized’, ‘fixed wing’, ‘runway’} → ‘routes include runways’ is a class. Practically, this is useful for planning the possible routes for the aircraft. Similarly, we can create the class (‘motorized’, ‘fixed wing’, ‘water’, ‘routes include body of water’) with the inference {‘motorized’, ‘fixed wing’, ‘water’} → ‘routes include body of water’.

Once these two classes are defined, another possibility arises. It might be possible to generalize the two classes into a higher level class by generalizing the properties used to define the inferences. In our example we can do this by creating two generalized properties: facility (with values ‘runway’ and ‘water’) and routes (with values ‘include runways’ and ‘include bodies of water’). The category “airplane” = {‘fixed wing’, ‘motorized’, facility, routes} with the inference {‘fixed wing’, ‘motorized’, facility} → {routes} is a superclass of the two new classes¹⁰. This is shown in Figure 6. For simplicity, we omit subclasses of “airplane” that reflect relationships between values of facility and routes.

¹⁰ To avoid clutter, we avoid using single quotation marks for these generalized properties (those having values) here and in the figures. In addition, unlike in the text, in the figures we do not use double quotation marks around class names.

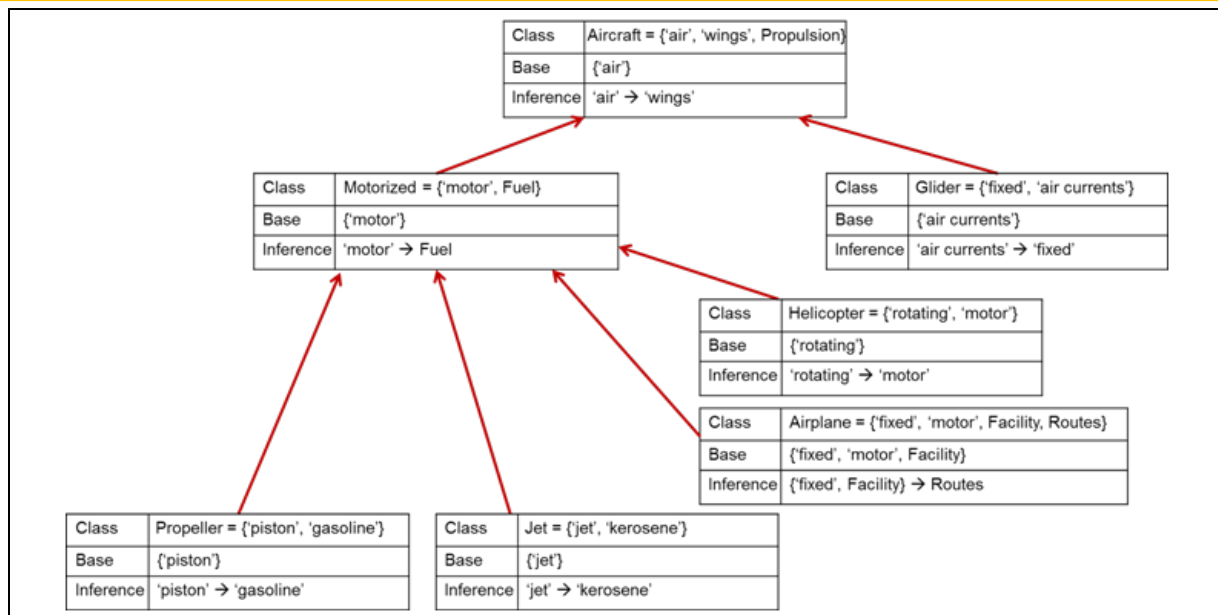
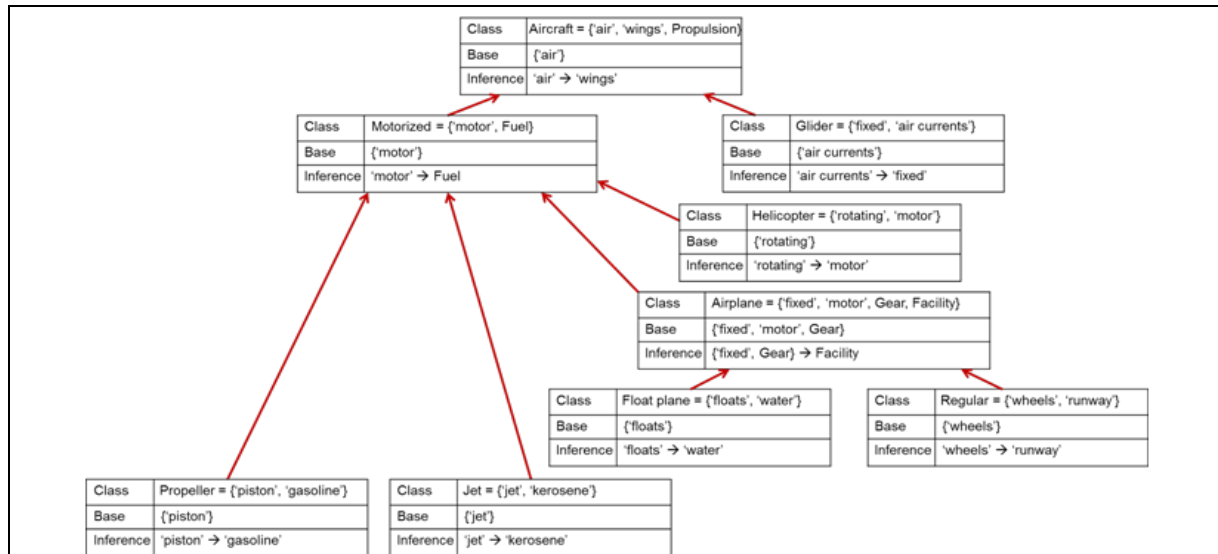


Figure 6. Aircraft Class Structure Reflecting Alternative Landing Facilities

Consider the second case: 'runway' becomes an inferred property and another property is found that is used (perhaps with other properties) to infer ("explain") the need for a runway. For example, the properties 'has wheels', 'has floats' can be used to infer the need for a runway or for a body of water as departure and landing facilities. As before, these properties can be generalized to landing gear (gear) and landing facility (facility), with an inference: gear → facility. The category "airplane" = {fixed, motorized, gear, facility} is a class with two subclasses: "regular plane" = {fixed, motorized, wheels, runway} and "float plane" = {fixed, motorized, floats, body of water}¹¹. Figure 7 shows this modified class structure.



This figure refers to the following relationships among attributes and attribute values (in addition to the ones in Figure 3): Gear: {floats, wheels}; Facility: {water, runway}.

Figure 7. Vehicle Class Structure with Inferences from Landing Gear to Facility

¹¹ This can be considered another way of applying the principle of maximal abstraction (as opposed to finding the minimal number of properties common to two classes). A possible advantage of this abstraction is that it points at the possibility of having more specific values for the generalized property.

The two possible modifications we could make to the class structure to maintain the information about runway can be compared to possible changes to scientific theories. In one case, some properties are used to “predict” a new property; in the other case, a new property is added to “explain” an existing property.

Table 3 summarizes the CIPS rules for dealing with contradictions to inferences.

Table 3. Rules To Modify a Class Structure when an Inference is Contradicted by Observation		
Use of rule	Principles	Rules/Guidelines
Always	Remove the inference	As a result a class may be removed. This will result in loss of properties.
Exploration: identify inferences	Consider a sub-category with the “lost” property	Does the sub-category have an inference that includes the “lost” property? If yes, this inference must include an additional property.
	Case 1: “Prediction”	The additional property is “predicted” by the “lost” property.*
	Case 2: “Explanation”	The additional property “explains” the “lost” property.*
Exploration: identify generalizations	If all instances in the original class now belong to new classes, abstraction might be possible to a superclass	The new superclass replaces the original “lost” class. The “lost” and additional properties are abstracted to a more generic property.
* Either prediction or explanation might “use” the other properties of the original class		

7.1. Comparing CIPS to Frames

Based on the above examples and analysis, we can summarize the differences between the CIPS model and frames. First, CIPS affords a structured approach to analyzing theories because: (1) CIPS comprises a homogeneous set of classes (all classes describe the same type of phenomena), and (2) every class must reflect inferences. These characteristics support analyzing exceptions and identifying possible modifications to a class structure (as demonstrated in the example above). The class structure provides a method to consider modifications in terms of creating new classes and asking under what conditions a category might be a class. In particular, the need to identify inferences and the principle of highest abstraction provide for a rule-based and procedural approach to theory testing and development.

In contrast, there is no requirement for frames to be homogeneous. There is no clear class structure, and different concepts (nodes) may refer to phenomena, to properties, or to property values. A requirement such as using the highest abstraction cannot be easily applied. Also, there is no requirement for frames to explicitly reflect inferences¹². The analysis of theories as described above (in the example of “aircraft”) depended on the distinction between categories and classes. This distinction is not reflected explicitly in the Frames approach. Additionally, the CIPS model supports the use of subclass and superclass relationships in reasoning about the domain. Thus, while the frames approach provides flexibility in describing phenomena, properties, and values, it does not support structured analysis of how theories can be modified to accommodate contradictory observations.

¹² Andersen et al. mention the importance of regularities: “empirical correlations between characteristics may be used to predict additional characteristics of an object that we recognize because it possesses some minimal set” (2006, p. 27). However, they claim that these relationships are captured in the ways objects are considered similar or not. They neither pose a requirement that useful categories possess such relationships, nor do they provide an explicit representation of inferences.

Second, the CIPS approach is scalable. Classes can be added simply by listing their properties, a base (a set of properties sufficient to identify membership), and the relevant inferences. Moreover, inferences can be specified explicitly for any type of property and, when applicable, in terms of formulae (calculating specific values of some properties from others). These characteristics of CIPS will be demonstrated below on an example of nuclear reactions. In addition, once classes are defined, the class structure (in terms of superclasses and subclasses) can be deduced. In contrast, when frames are used, sub-concepts are defined in terms of combinations of values of properties. Some inferences can be implied by the notion of contrast sets—categories defined by combinations of properties, where an instance can only conform to one combination. However, representing relationships among properties by indicating the categories that exist provides only a very limited, non-explicit representation of regularities. Moreover, if the number of relevant attributes or range of values for some attributes becomes large, the number of possibilities can make a frame-based representation prohibitively complex. Additionally, it might not be possible to represent regularities when property values are taken from a continuous set. Thus, the number of subclasses, their hierarchy, and the relationships among properties that can be represented in frames are limited.

Third, the CIPS approach is expressive. Abstractions in terms of superclasses and subclasses, and regularities (which reflect laws related to the classified phenomena), are represented explicitly. This is not the case for frames.

In summary, there are two types of differences between CIPS and frames. First, differences of representation enable CIPS to be scalable and to explicitly reflect regularities of a much wider variety than frames. Second, the principles of homogeneity, inference requirements, and maximal abstraction enable CIPS to provide a structured approach to reason about theory modification. Such an approach is not easily implementable using frames.

8. Applying CIPS in a Scientific Context

8.1. A CIPS View of Theories

We now suggest how the CIPS model can be used in a general scientific context. We begin with two observations. First, some scholars have suggested that classification is at the core of science. Andersen et al. (2006), referring to Kuhn's theory of scientific concepts, state: "According to this theory, the basic conceptual structure of science is a classification system that divides objects into groups according to similarity relations" (p. 20)¹³. This view is not limited to scientific inquiry that directly involves classifying the phenomena in a domain (such as the classification of celestial objects or biological organisms). Rather, it addresses organizing knowledge in science in general.

Second, consider the importance of regularities as manifested by relationships among observable properties of phenomena. In the natural sciences, such regularities are termed "laws". Laws are relationships among observable properties of phenomena¹⁴, such as Kepler's laws about planetary motion and Newton's law of gravity. Laws can have two origins. First, they can be empirical in nature; namely, they can emerge from (and generalize) observations—this was the case of Kepler's laws. Second, laws can be theoretical in their origin, as when theoretical analysis generates predicted relationships among observable properties¹⁵. This was the case for Newton's law of gravity¹⁶. To

¹³ Andersen et al. (2006, p. 20) proceed to state that: "The grouping is not determined by identifying necessary and sufficient conditions, but by learning to identify similarities and dissimilarities between the objects." However, for our purpose, what is important is the fundamental role of classification in science, independent of the mechanisms used to identify classes.

¹⁴ Hempel (1992) states: "The laws required for deductive-nomological explanations share a basic characteristic: they are, as we shall say, statements of universal form. Broadly speaking, a statement of this kind asserts a uniform connection between different phenomena or between different aspects of an empirical phenomenon" (p. 304).

¹⁵ Of course, an important motivation and measure of success for theory development is to derive the empirically found laws as consequences of the theory (Hempel, 1992, p. 304).

¹⁶ Hempel (1992) states: "This is not to say that 'empirical generalizations'—statements of universal form that are empirically well confirmed but have no basis in theory—never qualify as laws. Galileo's, Kepler's, and Boyle's laws, for example, were accepted as such before they received theoretical grounding. The relevance of theory is rather this: a statement of universal form, whether

summarize, science deals with identifying regularities among observable phenomena and seeking to explain them and potentially predict additional regularities. Regularities can be derived from observations (by induction) or from theoretical analysis (by deduction). A scientific law describes regularity identified in either way.

The notion of a scientific law in the above sense can be directly related to our model of classification. In a domain of phenomena, classes encapsulate both similarities (common properties) and regularities (inferences among properties). We therefore posit that, when representing scientific knowledge in classification terms, inferences reflect laws. The opposite should also hold: namely, the laws that are part of a scientific theory (whether empirical or theoretical) should be represented as inferences for classes in which the related phenomena are included. In other words, we suggest that scientific theories can be represented by class structures as defined in our model. This goes beyond simple categorization. Even if a category might appear “interesting”, to be useful it must be a class; that is, it must include inferences (conveying laws). Otherwise, the category is just shorthand for a set of similar phenomena¹⁷.

Moreover, beyond requiring that all categories included in a scientific theory should be classes, we also suggest that the principle of maximal abstraction can provide guidance where alternative ways exist to classify phenomena. This principle implies that if two classes of phenomena share some regularities (inferences), then a higher level abstraction—in terms of a class that manifests the common inference—should be formed. If, as a result, one of the original classes does not have additional inferences with respect to the new abstract class, then it loses its value as carrying useful knowledge (in terms of scientific laws) and should not be included in the class structure of the domain. In Section 8.2, we show how this principle can be used to modify theories when new observations contradict theoretical predictions. Before doing so, we first note several ways class structures and the CIPS model can be used in science. We consider three scenarios: applying a theory, testing a theory, and modifying a theory.

Applying a theory: An observed phenomenon can be matched with the classes of the theory. Inferences for all classes to which the phenomenon is matched can be used to predict additional properties (to those used for matching) of the phenomenon. Alternatively, it might be possible to identify the classes of interest to which the phenomenon might be matched. This can point to the need to perform more observations for some additional specific properties.

Testing a theory: Applying a theory requires sufficient information (properties) to identify classes to which a phenomenon can be assigned, or to determine whether or not the phenomenon belongs to specific classes. In contrast, testing a theory requires that, at least for one class, more properties will be observed than are necessary to identify the class. Some observed properties can then be used to identify classes to which the phenomenon belongs. Additional properties can be predicted based on the class inferences and compared to the properties actually observed. The inferences in this case are used as hypotheses to be tested. If a disagreement exists, provided there are no observation errors, the assumed inference (representing laws) does not hold.

Modifying a theory: In science, it is common when a new observation contradicts a theory (assuming no observational error) to modify the theory to accommodate the new observation. If we view scientific theories in terms of the CIPS model, this means the class structure needs to be modified. As indicated in introducing the CIPS model, this is akin to a feedback loop where the correction is made to the underlying class structure, not to the input (which is considered correct).

The ideas of good classification can help guide how a theory might be modified if a prediction is not corroborated. Clearly, the related law might be either: (1) invalid, or (2) limited in scope to only a subset of the category for which it was assumed valid before the new observation was made. In the

empirically confirmed or as yet untested, will qualify as a law if it is implied by an accepted theory (statements of this kind are often referred to as theoretical laws)...” (p. 306).

¹⁷ This does not diminish the role of categorization as the first step in identifying similar phenomena. This can be a step preceding identification of regularities.

first case, if this was the only inference distinguishing a class from other classes in the structure, the class will be eliminated. This indicates that what was considered a law of nature is not really so. In the second case, the question “what distinguishes the subclass for which the inference still holds?” arises. This question points out the need to identify additional properties of the related phenomena to distinguish instances of the new subclass from other instances of the original class. The search for such properties demonstrates how scientific questions can arise (and possibly lead to proposed empirical studies). As Section 7 discusses, once such a property is identified, two possibilities exist. First, the new property is inferred from (“predicted by”) the previously known ones. Second, the new property is used to infer (“explain”) previously known properties. Either possibility reflects a new law. It is also possible that, through abstraction of the properties involved in the new inference, a superclass of the new classes will be formed with an inference reflecting a more generalized law.

In Section 8.2, we demonstrate how CIPS can be applied to analyze changes in scientific knowledge using an example from the physics of nuclear reactions, one previously analyzed using frames (Andersen et al., 2006).

8.2. An Example: Nuclear Reactions¹⁸

Andersen et al. (2006) use a frame-based model to describe how prevailing theoretical predictions about nuclear reactions evolved to culminate in a major discovery—nuclear fission—at the time a radical shift in understanding the possible nuclear reactions physicists believed could occur. More specifically, the example relates to the reactions that can occur when a heavy nucleus is bombarded by neutrons (termed “projectiles”).

In the following, we apply CIPS to this example to:

- (1) Show how the CIPS model can be applied to scientific inquiry
- (2) Demonstrate how generic CIPS concepts can be instantiated in a specific domain
- (3) Show how to model discordances between empirical findings and scientific beliefs (formulated in terms of experimental and theoretical laws)
- (4) Show how CIPS-based reasoning can help resolve discordances, and
- (5) Demonstrate how the CIPS model can help identify questions to drive further research.

When neutrons were used to bombard a target made of heavy nuclei elements (such as uranium), several types of reactions were known to occur (based on experimentation):

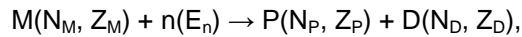
- (1) Alpha emission: The neutron reacts with the nucleus of a heavy element (the “mother” nucleus) and causes the emission of an alpha particle. An alpha particle (which is the nucleus of He⁴) has two protons and two neutrons. Hence, the resulting (“daughter”) nucleus has three fewer nucleons of which two are protons.
- (2) Proton emission: The neutron causes a proton to be ejected. The total number of nucleons in the daughter nucleus is the same as for the mother nucleus, and the number of protons decreases by one.
- (3) Neutron capture: The neutron is “captured”, leading to a nucleus with an additional neutron which decays by beta (electron) emission. The atomic mass and the atomic number (the number of protons) of the “daughter” nucleus (resulting after the beta emission) are each higher by one compared to the “mother” nucleus. While these are two consecutive reactions, we consider them as one for simplicity of exposition.

¹⁸ The descriptions in this section are somewhat incomplete for the sake of simplicity and so as not to include additional information to that of our comparison source (Andersen et al., 2006).

The theory in the early 1930s predicted that:

- (1) No particle heavier than an alpha particle will be emitted in these reactions, and
- (2) The possible reactions will depend on the projectile's energy.

In this example, the phenomena we seek to classify are reactions under neutron bombardment of heavy nuclei. We describe the reactions using the following notation (we do not follow here the common notation of physics):



where M is the mother nucleus, D is the daughter nucleus, and P is the emitted particle. N indicates the atomic number (total number of protons and neutrons, 0 for electron) and Z the charge (where the value for a proton is +1 and for an electron is -1), and E_n is the energy of the projectile neutron and can be 'l' (low or "slow") or 'h' (high or "fast").

Each phenomenon is described by seven properties: $\{N_M, Z_M, E_n, N_P, Z_P, N_D, Z_D\}$.

First we ask: is the category of reactions caused by neutron bombardment a class? The laws of physics state that the total number of nucleons (protons and neutrons) and the total charge do not change in these reactions. Hence, the following holds:

$$N_D = N_M + 1 - N_P \text{ and } Z_D = Z_M - Z_P$$

This implies that the properties describing the daughter nucleus can be inferred by knowing the properties of the mother nucleus and the emitted particle:

$$\{N_M, Z_M, N_P, Z_P\} \rightarrow \{N_D, Z_D\}$$

Note that the projectile energy, E_n , is part of the class definition but is not directly included in the inference. It follows that the category of neutron bombardment reactions $C^R = \{N_M, Z_M, E_n, N_P, Z_P, N_D, Z_D\}$ is a class.

As indicated above, the theory until 1938 predicted that:

- 1) No particle with $N > 4$ and $Z > 2$ (heavier or having more charge than an alpha particle) can be emitted under neutron bombardment
- 2) Neutron capture is much more likely for slow neutrons, and
- 3) An alpha particle can be emitted only when the projectile has high energy.

These three predictions can be described as the following inferences:

- 1) $N_P \neq 0$ or $Z_P \neq 0$ (any particle emitted) $\rightarrow N_P \leq 4$ $Z_P \leq 2$ and $|N_M - N_D| \leq 3$ and $|Z_M - Z_D| \leq 2$.

There are additional inferences about instances of the category C^R .

- 2) $Z_P = -1 \rightarrow E_n = 'l'$ (beta emission indicates that a slow neutron was captured)
- 3) $N_P = 4 \rightarrow E_n = 'h'$ (alpha emission indicates that a fast neutron was captured).

This can be written also: $E_n = 'l' \rightarrow N_P \leq 1, Z_P \leq 1$

Because there are three possible interaction types, three sub-categories of C^R are identified. Each is defined by restricting the values of the properties of the emitted particle (or, equivalently, of the daughter nucleus), bearing in mind that in our model values of properties are also properties:

C^A : Reactions leading to alpha emission $\{N_M, Z_M, E_n='h', N_P=4, Z_P=2, N_D, Z_D\}$

C^P : Reactions leading to proton emission $\{N_M, Z_M, E_n, N_P=1, Z_P=1, N_D, Z_D\}$

C^B : Reactions leading to beta emission $\{N_M, Z_M, E_n='l', N_P=0, Z_P=-1, N_D, Z_D\}$

We now consider whether each of these categories is a class; namely, whether it indicates inferences that are not generally correct for the class C^R (all neutron bombardment reactions). We first note that, for each category (of reactions), the daughter nucleus can be inferred by knowing the emitted particle. However, since this inference is included in the definition of the high level class C^R (of neutron bombardment reactions), it is not a new inference of the category.

Consider the category C^A (alpha emissions). Since an alpha particle can only be emitted for a high energy projectile, the property $E_n='h'$ is inferred. Therefore, C^A is a proper sub-class with the base $\{N_M, Z_M, N_P=4, Z_P=2\}$ and inferred property $E_n='h'$.

For the category C^P (proton emission), no similar inference holds. Hence, without additional inferred properties to those inferred for C^R , it is not a class.

Finally, for the category C^B (neutron capture) a low energy projectile is inferred. Hence, C^B is a proper sub-class with the base $\{N_M, Z_M, N_P=1, Z_P=1\}$ and inference $E_n='l'$. Additionally, another particle is emitted in beta emissions—a neutrino (this is not part of the example in Andersen et al.). With this addition, C^B has two inferences.

So far, we have identified the class structure that describes the pre-1936 knowledge:

Neutron reactions: $C^R = \{N_M, Z_M, E_n, N_P, Z_P, N_D, Z_D\}$ ¹⁹
 Base = $\{N_M, Z_M, E_n, N_P, Z_P\}$ Inferences: $N_P \leq 4; Z_P \leq 2; N_D = N_M + 1 - N_P; Z_D = Z_M - Z_P$

Alpha emission: $C^A = \{N_M, Z_M, E_n='h', N_P=4, Z_P=2, N_D, Z_D\}$
 Base = $\{N_M, Z_M, N_P=4, Z_P=2\}$ Inference: $E_n='h'$

Neutron capture (Beta emission) C^B : $\{N_M, Z_M, E_n, N_P=0, Z_P=-1, N_D, Z_D, (\text{neutrino})\}$
 Base = $\{N_M, Z_M, N_P=0, Z_P=-1\}$ Inferences: $E_n='l'$; (and neutrino).

If no additional properties or inferences distinguish proton emission reactions, this structure suffices to describe the pre-1936 knowledge (Figure 8).

Andersen et al. (2006) use frames to model three shifts, each reflecting a new empirical discovery that occurred in nuclear physics during the 1930s. The discoveries were: low energy alpha emission, neutron chipping, and nuclear fission. Of these, nuclear fission was the most radical one. We now discuss how the first and third of these shifts can be analyzed using the CIPS model.

¹⁹ This class is defined in terms of general properties. For specific phenomena, these properties will acquire specific values (e.g. number of nucleons or of protons). The inferences here relate values of the more general properties.

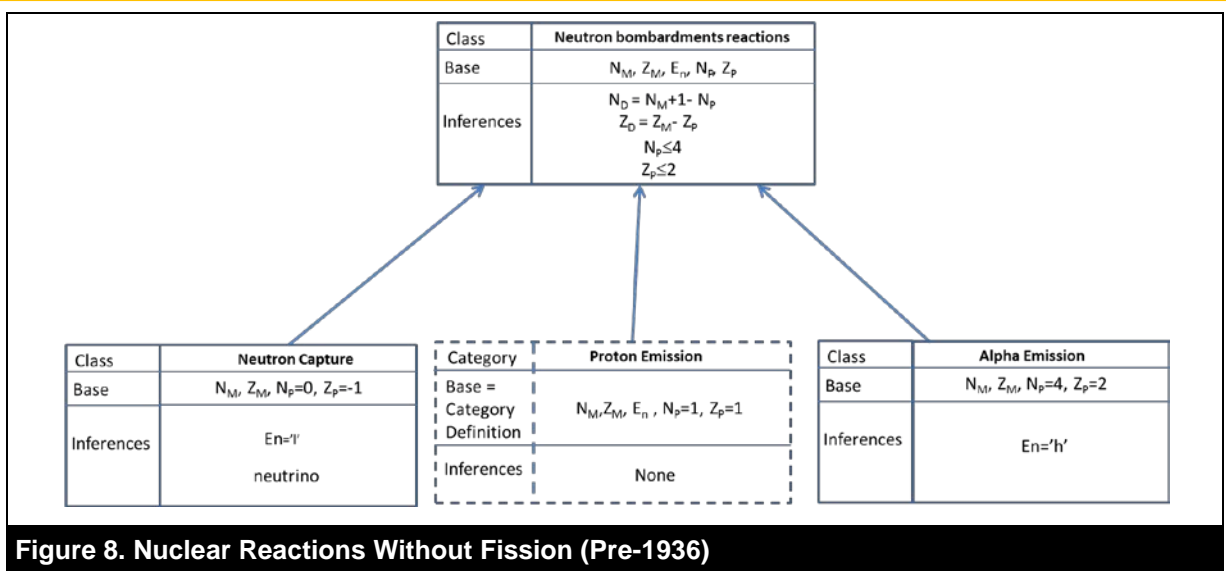


Figure 8. Nuclear Reactions Without Fission (Pre-1936)

First, in 1936, it was discovered that an alpha particle can be emitted when uranium-238 was bombarded by low energy neutrons ($E_n='l'$). Consider now an observed alpha emission $\{E_n='l', N_P=4, Z_P=2\}$. In CIPS, this observation would match classes C^R and C^A , and the inferences would be:

- 1) For C^R : $N_D=235, Z_D=90$
- 2) For C^A : $E_n='h'$.

There clearly is a contradiction for E_n . When this was discovered, the suggestion at the time was to consider this a special case possible only for some mother nuclei. In our terminology, this means adding a category: $C = \{N_M, Z_M=92, E_n='l', N_P=4, Z_P=2, N_D, Z_D\}$. Moreover, if the inference of C^A (alpha emission) related to E_n does not hold anymore, C^A would not be a class unless it has at least one additional inference to those of its superclass C^R . Such an inference would require at least one new property.

However, an alternative might exist. As now there are two possible categories of phenomena related to alpha-emission ($E_n='l', E_n='h'$), there might be additional properties and inferences that indicate when the reaction can happen with low energy and when it can happen only with high energy. Such properties and inferences would reflect the underlying reasons (in terms of physical laws) why such a reaction is possible for some nuclei and not for others. This demonstrates how further scientific questions can be generated in the CIPS framework.

The third and most radical change occurred in 1938 when the chemical properties of the generated elements for a uranium target were studied more carefully. It was found that one of the daughter nuclei was barium ($Z=56$). When comparing to the inferences of the class C^R (all neutron bombardment reactions), this contradicts the inferences $N_P \leq 4; Z_P \leq 2$. In our model this will require two changes:

- 1) The inference $N_P \leq 4; Z_P \leq 2$ does not hold anymore, and
- 2) A new category is added to describe fission reactions:
 $C^F = \{N_M, Z_M, E_n, N_P > 4, Z_P > 2, N_D, Z_D\}$.

In this notation, we consider one of the fission products the emitted particle.

These changes raise two questions. First, the inference that limits N_P and Z_P is not valid anymore. It was part of the definition of the class C^R (all neutron bombardment reactions). Hence, the question arises whether or not this category is still a class. Since the inference related to daughter nuclei is still valid, it is still a class. Second, the category C^F can be considered a class if and only if it provides some inferences not included in its superclass C^R . This points out to the need to find out if the fission reaction has additional properties and inferences related to them. The answer is positive. A fission reaction is accompanied by a very high energy release (much higher than the projectile's energy or the energy released in any of the previously known reactions), and by emission of neutrons. It is exactly these properties that make the fission reaction of much interest because they enable both nuclear reactors and nuclear fission weapons. It follows that with the properties 'very high energy released' and 'neutrons emitted' added to the description of the classified phenomena, the category "fission reactions" is a class:

Fission:

$C^F = \{N_M, Z_M, E_n, N_P > 4, Z_P > 2, N_D, Z_D, \text{'very high energy released'}, \text{'neutrons emitted'}\}$, with base $\{N_M, Z_M, E_n, N_P > 4, Z_P > 2\}$ and inferred properties 'very high energy released' and 'neutrons emitted'.

The new class structure (Figure 9) is now:

Neutron Reactions: $C^R = \{N_M, Z_M, E_n, N_P, Z_P, N_D, Z_D\}$

Base = $\{N_M, Z_M, E_n, N_P, Z_P\}$ Inferences: $N_D = N_M + 1 - N_P$; $Z_D = Z_M - Z_P$

(here N_P is the sum of the nucleons in one of the fission products and the number of neutrons emitted, and N_D is the number of nucleons in the other fission product)

Alpha emission: $C^A = \{N_M, Z_M, E_n = 'h', N_P = 4, Z_P = 2, N_D, Z_D\}$

Base = $\{N_M, Z_M, N_P = 4, Z_P = 2\}$ Inference: $E_n = 'h'$

Neutron capture (beta emission): $C^B = \{N_M, Z_M, E_n, N_P = 0, Z_P = -1, N_D, Z_D, (\text{neutrino})\}$

Base = $\{N_M, Z_M, N_P = 0, Z_P = -1\}$ Inferences: $E_n = 'l'$; (neutrino)

Fission:

$C^F = \{N_M, Z_M, E_n, N_P > 4, Z_P > 2, N_D, Z_D, \text{'very high energy released'}, \text{'Neutrons emitted'}\}$

Base = $\{N_M, Z_M, E_n, N_P > 4, Z_P > 2\}$

Inferences: 'very high energy released', 'Neutrons emitted'.

Finally, to demonstrate how CIPS can invoke or be used to frame scientific questions, we mention two examples that can be motivated by the class structure of Figure 9. First, there is a category of reactions—proton emission—that seems to have no inference that would reflect an underlying physical law. It would make sense to ask what could be the properties of this reaction that would be involved in inferences for this category.

Second, of the three subclasses of reactions in Figure 9, two have inferences related to the projectile energy. This points to the question: "Is there any inference about fission reactions related to the energy of the impinging neutron (E_n)?" The answer might be given by identifying some additional properties (or property values) involved in such inference that would reflect a physical law. Indeed, such a law was found by Bohr in 1939, and it was based on the number of neutrons and protons in the mother nucleus. In particular, isotopes with very high atomic number were predicted to fission when bombarded by slow (low energy) neutrons if the number of their neutrons was odd²⁰.

²⁰ The reason is that neutrons (and protons) tend to form pairs as this is a lower energy state. When bombarded with a neutron, an unpaired neutron in the nucleus would have a neutron to pair with, leading to a release of energy sufficient to split the nucleus (for heavy nuclei).

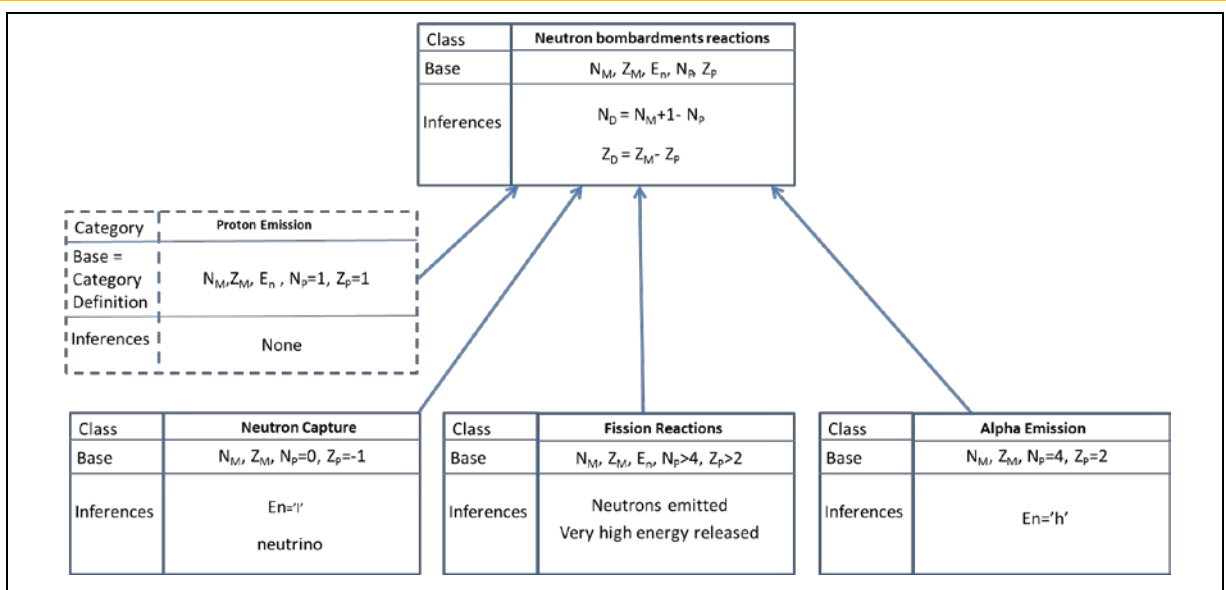


Figure 9. Nuclear Reactions With Fission

9. Conclusion

In this paper, we suggest a model of categorization and classification based on the role of efficiency and inference in cognition and on the need to balance the effort and time required to acquire, maintain, and infer information about phenomena. A critical contribution of CIPS is the distinction between categories and classes, where classes are categories that include inferences. In CIPS, information is processed by combining acquisition via observation and communication, use of prior knowledge of individual phenomena, class matching, and inference. Usually, many choices of classes and class collections are possible. The ones chosen should be determined by the purpose of the classification system, capturing the inferences of interest (as the example of bodies in our solar system demonstrated) and by aiming at the highest possible abstractions and minimal redundancies.

The CIPS model was originally developed in an effort to create principles to guide classification in the information systems domain and to design mechanisms to ensure that classes chosen for an application meet some criteria of “goodness” (Parsons, 1996; Parsons & Wand, 1997, 2008b). In this paper, we extend the use of the principles to other fields of research. For example, distinguishing the concepts of category and class helps in understanding why the debate over the demotion of Pluto from planetary status is a non-issue (Parsons & Wand, 2008a). The CIPS model indicates that what is important is that any structure used in this domain consists of classes (rather than categories). Identifying classes is at the core of scientific progress.

We believe the CIPS framework can be used to suggest deep analogies between completely different types of systems (Holland et al., 1984), such as technological information systems, medical diagnosis and the immune system (see Appendix).

Beyond this, we suggest that the CIPS model can be used to represent and analyze scientific knowledge in general, and compare it to an existing approach for such analysis—frames (Andersen et al., 2006). Specifically, we propose that a scientific theory can be represented as a class structure. We show by example (nuclear reactions) the use of CIPS to analyze the application, testing, and modification of scientific theories. Of critical importance is the distinction between categories and classes. Classes must have inferences that relate the properties of their instances. When using CIPS to describe a scientific theory, these inferences represent scientific laws.

In summary, we propose that the CIPS model is useful for: (1) providing a common set of concepts

and principles to analyze classification-related phenomena in disparate domains, both artificial and natural; (2) gaining understanding into the nature of controversies about classifications, (3) framing scientific knowledge and theories in terms of class structures, (4) formalizing discordances between empirical findings and prevailing scientific beliefs (both experimental laws and theories), and (5) contributing to the resolution of discordances by identifying questions that can drive further research.

We note two possible extensions of CIPS. First, membership in a category may be approximate or fuzzy (e.g., being “tall”) rather than binary (“belongs” or “does not belong”). This would not affect the general view of a CIPS but might impact the specific mechanisms of pattern matching and inferences, and can lead to “approximate” inferred properties. Second, the choice of a class structure should reflect the benefits and costs of the collection. The above guidelines for collections entail (via maximal abstraction and minimal redundancy) a naïve assumption that benefits and costs are directly related to the number of properties and classes. In reality, the benefits that can be attributed to inferred properties and the effort needed to observe or maintain some information will likely depend on the nature of the information (in terms of properties of phenomena) and its uses. Thus, much more detailed analysis may be done regarding how collections of classes should be determined. The considerations can include time, storage capacity, and effort (resources) needed for acquisition (by observation or by communication), storage and retrieval (of instance properties and class patterns), pattern matching, and performing inferences. Since classification is driven by the need to economize these resources, the actual classes that arise would be the outcome of “economizing” and “balancing” of resources. Again, we believe that such efforts can inform many diverse and seemingly completely different domains.

Acknowledgements

We are grateful to the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada for supporting this research. We also thank the Senior Editors of the special issue and the reviewers for detailed feedback that substantially improved the paper.

References

- Andersen, H., Barker, H. P., & Chen, X. (2006). *The cognitive structure of scientific revolution*. Cambridge: Cambridge University Press.
- Bowker, G., & Starr, S. L. (2000). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.
- Carter, J. H. (2000). The immune system as a model for pattern recognition and classification. *Journal of the American Medical Informatics Association*, 7(1), 28-41.
- Chandrasekaran, B., Josephson, J., & Benjamins, R. (1999). What are ontologies, and why do we need them? *IEEE Intelligent Systems*, 14(1), 20-26.
- Dickinson, J. L., Zuckerberg, B., & Bonter, D. N. (2010). Citizen science as an ecological research tool: Challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, 41, 112-149.
- Forrest, S., & Hofmeyr, S. A. (2000). Immunology as information processing. In L. A. Segel and I. R. Cohen (Eds.), *Design principles for immune system & other distributed autonomous systems* (pp. 361-387). New York: Oxford University Press.
- Hempel, C. (1992). Laws and their role in scientific explanation. In R. Boyd, P. Gasper, and J. D. Trout (Eds.), *The philosophy of science*. Cambridge, MA: MIT Press, 299-315.
- Hirschheim, R., Klein, H. K., & Lyytinen, K. (1995). *Information systems development and data modeling*. Cambridge: Cambridge University Press.
- Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction*. Cambridge, MA: The MIT Press.
- IAU. (2006, August). The Final IAU Resolution on the definition of "planet" ready for voting. Retrieved April 29, 2013, from http://www.iau.org/public_press/news/detail/iau0602/
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Lukyanenko, R., Parsons, J., & Wiersma, Y. F. (2011). Citizen science 2.0: Data management principles to harness the power of the crowd. *Proceedings of the Conference on Design Science Research in Information Systems and Technology, Milwaukee, WI*, 465-473.
- Parsons, J. (1996). An information model based on classification theory. *Management Science*, 42(10), 1437-1453.
- Parsons, J., & Wand, Y. (1997). Choosing classes in conceptual modeling. *Communications of the ACM*, 40(6), 63-69.
- Parsons, J., & Wand, Y. (2000). Emancipating instances from the tyranny of classes in information modeling. *ACM Transactions on Database Systems*, 25(2), 228-268.
- Parsons, J., & Wand, Y. (2003). Attribute-based semantic reconciliation of multiple data sources. *Journal on Data Semantics*, 1, 21-47.
- Parsons, J., & Wand, Y. (2008a). A question of class. *Nature*, 455(7216), 1040-1041.
- Parsons, J., & Wand, Y. (2008b). Using cognitive principles to guide classification in information systems modeling. *MIS Quarterly*, 32(4), 839-868.
- Parsons, J., Lukyanenko, R., & Wiersma, Y. F. (2011). Easier citizen science is better. *Nature*, 471, 37.
- Rehder, B., & Burnett, R. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology*, 50, 264-314.
- Ritvo, H. (1997). *The platypus and the mermaid and other figments of the classifying imagination*. Cambridge, MA: Harvard University Press.
- Rosch, E. (1978). Principles of categorization. In E. Rosch and B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.
- Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology and Evolution*, 24, 467-471.
- Smith, E. (1988). Concepts and thoughts. In R. Sternberg and E. Smith (Eds.), *The psychology of human thought*. Cambridge, MA: Cambridge University Press.
- Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142, 2282-2292.
- Wiersma, Y. F. (2010). Birding 2.0: Citizen science and effective monitoring in the Web 2.0 world. *Avian Conservation and Ecology*, 5(2), 13.
- Yamauchi, T., & Markman, A. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 776-795.

Appendix. CIPS and Biological Systems

Below we discuss two biological systems to demonstrate briefly how CIPS can support analysis of phenomena in other domains or adaptation of ideas from other domains²¹.

The Immune System

The literature has recognized information processing aspects of the immune system, independent of biological processes (Carter, 2000; Forrest & Hofmeyr, 2000) and applied this view to analyze immune-related phenomena and to inform research on computer systems (Forrest & Hofmeyr, 2000). Forrest and Hofmeyr state: “[T]he success of all analogies between computing and living systems will ultimately rest on our ability to identify the correct level of abstraction” (2000, p. 382). We propose that the CIPS view can provide this level of analysis. Two examples follow. First, the immune system “must be able to recognize foreign material (“antigens”) as foreign” and to do this, “the immune system has recognized ... patterns ...” (Forrest & Hofmeyr, 2000, pp. 362-363). In CIPS, “input” will be the biological characteristics of cells to be recognized (antigens) and the patterns can be considered class definitions (stored in antibodies). Second, CIPS predicts that an input phenomenon can be matched to several classes. Forrest and Hofmeyr (2000) state that “receptors do not require an exact match to an antigen in order to be activated” and that in immunology this leads to cross-reactivity which might be beneficial or harmful, depending on whether the identified classes correspond to pathogens, or to self cells, respectively.

The CIPS view can help import ideas from the immune system to IT artifacts. Four examples follow. First, in the immune system memory and processing are completely distributed and performed by independent units—immune system cells—where different types of cells interact (e.g., T cells and B cells (Carter, 2000, p. 29)). This suggests a possible CIPS architecture of a distributed repository of class definitions and execution of inferences. An input phenomenon can be then compared concurrently to various class definitions that can be stored in disparate locations. Second, antibodies in the immune system “decay” over time. Similar mechanisms can be adapted by IT-based systems to, over time, eliminate classes that are not used. Third, the immune system uses fast replication of cells following recognition of a foreign organism. In CIPS, this could amount to “spawning” more processing units representing classes that match inputs. Such an approach can improve processing speed. Fourth, the immune system uses rapid mutation and two selection processes in tandem (Forrest & Hofmeyr, 2000; Carter, 2000). Positive selection leaves cells that can recognize foreign proteins. Negative selection eliminates cells that recognize self-proteins. In CIPS, similar mechanisms can support automatic evolution of classification schemes, a common objective of both system design and of business intelligence and “big data” applications.

The Amygdala

The amygdala is a brain structure that performs very fast recognition (and inference) based on partial input. Such processing can be very fast, but lead to a high (“crude”) level of abstraction, which might call for additional processing to determine required actions. This suggests a two-step process in CIPS. First, a very fast identification of classes can generate a set of “candidate” inferences. Second, additional processing (using additional information), can be applied to choose among possible inferences. This can support “focusing” processing resources on more “promising” possibilities (e.g. in “Big Data” applications).

²¹ A more detailed analysis is available from the authors.

About The Authors

Jeffrey PARSONS is University Research Professor and Professor of Information Systems in the Faculty of Business Administration at Memorial University of Newfoundland. He holds a Ph.D. in Information Systems from the University of British Columbia. His research interests include conceptual modeling, data management, business intelligence, and recommender systems; he is especially interested in classification issues in these and other domains. His research has been published in journals such as *Nature*, *Management Science*, *MIS Quarterly*, *Information Systems Research*, *Journal of Management Information Systems*, *Communications of the ACM*, *ACM Transactions on Database Systems*, and *IEEE Transactions on Software Engineering*. Jeffrey Parsons has served in several editorial roles, both for journals and conferences.

Yair WAND is CANFOR Professor of MIS at the Sauder School of Business, The University of British Columbia, Canada. He received his D.Sc. in Operations Research from The Technion (Israel Institute of Technology), his M.Sc. in Physics from the Weizmann Institute (Israel), and his B.Sc. in Physics from the Hebrew University, Jerusalem. Yair's research interests focus on theoretical foundations and methods for information systems analysis and design. In particular, he has done work on ontological approaches to information systems, on theoretical and empirical methods to study conceptual modeling, and on the application of classification principles in modelling and design of information systems.