

ESSAY

A question of class

Fundamental misunderstandings about classification can lead scientists down unproductive or dangerous paths, argue **Jeffrey Parsons** and **Yair Wand**.

Consider this list of scientific controversies: the decision by the International Astronomical Union to define 'planet' in such a way as to exclude Pluto; the ongoing debate over what 'species' means for bacteria; and the discovery, initially controversial, that the most common cause for peptic ulcers is an infection by the *Helicobacter pylori* bacterium. What do these diverse issues have in common? They are all examples of scientists disagreeing over how to classify phenomena 'correctly'.

Categorization is a fundamental skill learned in childhood. Yet the principles that guide it are sometime misunderstood and misused by scientists. Here, we analyse the concepts of 'category' and 'class', and reveal that some controversies over scientific classification, such as the case of planets, should not be controversial at all. Others, including the case of ulcers, can be explained as a consequence of wrong assumptions about categories.

Efforts to advance knowledge that is based on ill-conceived classifications can prove futile, and even harmful. At best, they might result in wasted time spent arguing over terminology. More seriously, they can misdirect research efforts and funding. And at worst, in cases such as the misclassification of medical conditions, the result can be serious harm, including misdiagnosis, improper treatment and even death.

To avoid such consequences, the scientific community should recognize that classification is a purposeful human activity that reflects observations about relationships among properties of phenomena. As a variety of relationships occur in nature, different classification schemes can coexist. On the other hand, classification that does not reflect true relationships can misguide scientific discourse.

Classification has long been a subject of research in cognitive psychology, and it is recognized as an evolved mechanism that supports survival. Being able to categorize actions and objects to anticipate possible 'good' or 'bad' outcomes is a critical skill for humans. At a basic level such predictions help us acquire food and shelter, or keep us from being eaten.

In science, classification is rarely related to survival so directly, yet the underlying principles are the same. Phenomena can be categorized in



many ways on the basis of the properties they share. However, categories are useful only if they make it possible to infer further information, and only if they do so consistently and over a reasonable time period. To distinguish a general category from a more useful one with inferences, we call the latter a 'class'.

Whereas a category simply reflects a repeating pattern of properties, a class additionally indicates that relationships exist between these properties, even if the mechanisms behind the relationships are unknown. For example, all the things in a room make up a category (they share the property of being in the room), but not necessarily a class; their presence in the room might not reflect some deeper underlying principle. In contrast, 'sea mammals' have the properties 'live in the sea', 'can regulate their body temperature within a narrow range', 'breathe air' and 'have blubber'. Having the first two properties is sufficient to recognize a creature as a sea mammal; the other properties can then be inferred reliably (indicating

some underlying relationships between the properties). Hence, 'sea mammals' is a class.

Many scientific discoveries begin with the identification of repeating patterns, leading to the formation of categories. If further discovery indicates that relationships exist between the properties of a category (that is, the group forms a class), this is often the first hint of some underlying law of nature. In structuring knowledge, scientists should aim to identify classes rather than categories, and controversies over classification should be understood and resolved in that context.

Planetary prowess

In 2006, the International Astronomical Union approved the following definition: "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." It turns out that all

D. PARKINS

celestial bodies that have these properties also have non-intersecting orbits and lie close to the ecliptic, giving this defined concept predictive power and making the definition a class. Pluto, however, does not have property (c) or the two additional predicted properties.

In our view, the fact that the International Astronomical Union definition of planet constitutes a class probably indicates a deep similarity between these bodies that is not shared by Pluto. Yet there are other ways of defining classes of bodies in the Solar System that will include the planets (according to the new definition) and Pluto, and might prove useful to scientists interested in other factors. A definition based on properties (a) and (b) alone, for example, includes Pluto, as well many other similar bodies (Plutinos). All these bodies share at least one other property — they are not large enough to have ever sustained thermonuclear reactions. Hence, this definition forms a distinct (but unnamed) class, the membership of which overlaps that of ‘planet’.

Classes of bodies in our Solar System can be defined in many other ways, reflecting interrelationships associated with how the bodies were formed, their physical characteristics (size, composition, shape), their dynamic characteristics (orbit, rotation), or even whether they have an environment that can support life. Each of these definitions is useful for different reasons. As a result, supporting a multitude of definitions would be useful to the community and, in fact, this position has been taken by some astronomers. Thus, the controversy over a sole correct definition of ‘planet’, and whether Pluto falls within it, is unwarranted from a scientific perspective. Instead, it simply reflects historical and emotional associations with a specific term — ‘planet’.

Although the Pluto controversy involves the classification of a specific object, in biology an entire classification system is in doubt — as evolutionary biologist Ernst Mayr wrote in his book *What Evolution Is* (Basic Books, 2001), “naturalists have had a terrible time trying to reach a consensus” on the notion of species. In particular, he added, “there is considerable uncertainty of how many ‘species’ of bacteria to recognize”.

There are conflicting objectives in categorizing life forms. The more shared properties per category, the greater the potential predictive power, making highly specialized categories attractive. For example, knowing that an animal is a marsupial implies more than knowing it is a mammal, which implies more than knowing it is a vertebrate. But very specialized categories are not stable over time. Members of different subspecies, for example, can interbreed, such that a subspecies can lose

its predictable, distinct (from other subspecies) set of properties. Thus, an optimal level between predictive power and stability needs to be found.

For sexually reproducing organisms, the categorization level of ‘species’ delivers this compromise. But the case is much less clear for other organisms. For bacteria and archaea, extensive lateral gene transfer has placed in doubt even the idea of hierarchical classification. The rapid swapping of genes between these life forms means that their properties are not stable over time. When trying to construct a family tree, organisms that belong to different branches can share so many similarities that the tree turns into something more like a net. The identification of core genes has been helpful in categorizing certain microbes, but these organisms can easily acquire non-core genes that change their properties. Some researchers have controversially posited that the constant changeability of bacteria makes ‘species’ a meaningless label.

Much work in the microbiology community has been focused on these issues. Instead of asking ‘what is a species?’, biologists need to ask the more fundamental question ‘what are the useful, stable, related properties that can be used to define a class?’. The answer for microbes is not clear. Perhaps it lies in finding maximal groups of coexisting sets of genes that persist within certain environments. Perhaps, as with planets, there are multiple solutions that allow for useful predictions, depending on the circumstances. The important thing is to avoid the trap of imposing a categorization solution that fits some organisms onto others that it does not fit.

Defining disease

In the case of disease, one of the most important classifications used in diagnosis is that of aetiology. Based on causation, diseases can be placed in one of three categories: genetic, environmental or pathogenic. These categories are also classes because they imply additional information, such as possible prognosis and effect of treatments.

In the 1950s, ulcers were placed quite firmly in the class of environmentally caused diseases rather than those caused by a pathogen. Variable delays between infection and onset of ulcers, and the difficulty of growing suspect bacteria *in vitro*, led to a widespread belief that bacteria could not live in the acidic environment of the stomach. Stress and diet were instead thought to be causes. Although

even in the 1940s there had been indications that peptic ulcers could be cured by antibiotics, the overriding assumption that the condition was environmental blinded scientists and doctors to the implications of those findings. This incorrect classification of a medical condition not only hindered the discovery of the causative factor, but also delayed its acceptance. Even though *H. pylori* was strongly implicated as a possible major cause of ulcers in 1982, it was not until 1994 that antibiotics were generally recommended for their treatment, and as late as 1995 only 5% of patients with ulcers were receiving antibiotic treatment.

When considering the reasons why the bacterial hypothesis was missed for such a long time (and then not readily accepted), the main problem was the misattribution of the property ‘cannot grow in the acidity of the stomach’ to the class of bacteria. Re-evaluating this fundamental property involved a major mind-shift that was difficult to accept.

Taking a classification perspective on scientific discourse suggests a sequence of questions to ask when studying a domain of phenomena. What are the properties of interest of these phenomena? Are there stable sets of properties common to these phenomena? Are there stable relationships in some of these sets? And finally, and most importantly, what is the evidence or rationale that these relationships reflect the true nature of the phenomena? This perspective has two important implications. First, scientists should make every effort to ensure that the assumed relationships among properties are indeed correct. Second, rather than arguing over which of several classification schemes is preferable, researchers should recognize that several correct and useful schemes can coexist. And overall, scientists should recognize that classification happens in the mind and, as a result, it can be influenced by beliefs and emotions. This is where science can go astray. ■

Jeffrey Parsons and **Yair Wand** are professors of information systems at, respectively, Memorial University of Newfoundland and the University of British Columbia.
e-mails: jeffreyp@mun.ca, yair.wand@ubc.ca

See <http://tinyurl.com/3lubs0> for further reading. See also News Feature, page 1023. Discuss definitions online at <http://tinyurl.com/4afapl>.

Correction

In the Essay “Beijing 1987: China’s coming-out party” (*Nature* **455**, 598–599; 2008), we misspelt the name of Julia Marton-Lefèvre; this has been corrected in the online version.

“The controversy over a sole correct definition of ‘planet’, and whether Pluto falls within it, is unwarranted from a scientific perspective.”