Concept Utility

Paul Égré Cathal Ó Madagáin

Abstract

Practices of concept-revision among scientists seem to indicate that concepts can be improved. In 2006, the International Astronomical Union revised the concept PLANET so that it excluded Pluto, insisting that the result was an improvement. But what could it mean for one concept or conceptual scheme to be better than another? Here we draw on the theory of epistemic utility to address this question. We show how the plausibility and informativeness of beliefs, two features that contribute to their utility, have direct correlates in our concepts. These are how *inclusive* a concept is, or how many objects in an environment it applies to, and how *homogeneous* it is, or how similar the objects that fall under the concept are. We provide ways to measure these values, and argue that in combination they can provide us with a single principle of *concept utility.* The resulting principle can be used to decide how best to categorize an environment, and can rationalize practices of concept revision.

Keywords: epistemic utility; concepts; concept revision; conceptual change; homogeneity; inclusiveness; informativeness; plausibility;

Introduction

Some ways of conceptualizing our environment seem obviously better than others. Given our astronomical knowledge, it seems right to group the objects in the solar system under the categories PLANET, STAR, ASTEROID, and MOON.[1] We would be reluctant, on the other hand, to group them into the three categories STAR OR ASTEROID, SMALL PLANET, and MOON OR LARGE PLANET. Even though the second set of categories covers the same objects as the first, it is intuitively a poor conceptualization of the domain. Similarly, scientists often come to the conclusion that a conceptual scheme they are working with could be improved, sometimes in light of new discoveries. A recent example is a 2006 resolution of the International Astronomical Union. Here it was decided that in light of the discovery of a range of celestial objects in the vicinity of Pluto, it would improve our categorization of the solar system if we excluded Pluto from the concept PLANET

[1]Throughout we use small caps to denote concepts or categories, and we use the terms 'concept' and 'category' interchangeably.

and grouped it instead with the new objects. But what does it mean for one concept or conceptual scheme to be better than another?

Here we approach this question from the perspective of epistemic utility theory, a branch of epistemology that aims to describe what makes our beliefs useful (e.g. Carnap 1928, Popper 1955, Levi 1967, Maher 1993, Huber 2007). In this literature, two elements of beliefs are widely regarded as fundamental to their utility: their plausibility, or how likely they are to be true, and their informativeness, or how much they tell us about the world. We show here that concepts, as the components of beliefs, have two properties that directly correlate with the plausibility and informativeness of beliefs that we form in using those concepts. These are the *homogeneity* of a concept, which contributes to the plausibility of inductive generalizations we make using that concept; and the *inclusiveness* of a concept, which contributes to the informativeness of those generalizations. We provide measures for these aspects of concepts, and argue that in combination they provide us with a principle of *concept utility*. The resulting principle allows us to directly compare the utility of competing conceptual schemes, and to explain and rationalize practices of concept revision.

Several proposals already exist for how to go about categorizing an environment, but they exhibit substantial shortcomings. After briefly reviewing these proposals (section 1), we turn to the question of epistemic utility, and the value of both plausibility and informativeness to belief (section 2). We then explore how conceptual schemes can support these values, and introduce our principle of concept utility (section 3). We illustrate how the principle allows us to determine and revise a conceptual scheme with a toy example (section 4), and finally we show how our account explains a real case of concept revision, in the puzzling reclassification of Pluto (section 5).

1 Ways of Conceptualizing

One way to conceptualize an environment is to group its elements together in terms of similarity – something proposed by Carnap (1928), and reflected in accounts of concepts such as the prototype theory (Rosch and Mervis 1975), the exemplar theory (Gopnik and Meltzof 1997) and the conceptual spaces account (Gärdenfors 2000). Certainly, grouping according to similarity would allow us to decide between the intuitive categorization of the solar system described above, and the odd categorization considered next to it – the categories in the first set are more internally similar. However, it is doubtful that similarity is the only thing that drives our categorization practices. Due to convergent evolution, dolphins and sharks are very similar – much more similar than dolphins and dogs, for example. And yet, we categorize dolphins with dogs, as mammals, and not with sharks. Equally, if we paint a racoon to look like a skunk, and even physically alter it to have the capacity to produce foul smelling spray from a gland in its rear, children from as young as 6 years will nevertheless insist that it is still a racoon (Keil 1989). As we shall see, similarity is only partly what we are concerned with when we categorize.

Another way to conceptualize an environment is to try to divide it in terms of 'natural kinds', something 'externalist' theories often appeal to (Putnam 1975, Kripke 1980, see also the 'theory-theory', Keil 1989).[2] This approach need not depend on similarity, since members of the same natural kind can fail to resemble one another. Unfortunately it is rarely easy to decide what counts as the most natural way to categorize an environment. Often multiple mutually exclusive ways of categorizing the same environment can have legitimate claims to being natural (Dupré 1993). There are many conflicting ways to distinguish species in natural terms, for example. On an 'interbreeding' account (Mayr 1969), species are populations that can reproduce with one another and not with members of other groups. On a 'cladistic' account, species are distinguished by ancestry (Cracraft 1983). However, some cladistically grouped species cannot interbreed, such as species with a common ancestor that reproduce asexually. These will be grouped together by a cladistic account, but separately by the interbreeding account (cf. Ereshevsky 1998). Since there are many conflicting ways to decide which things belong together naturally, the value of a conceptual scheme cannot be simply decided by the question whether it is natural.

A further way to decide how to conceptualize is to consider the impact a conceptual scheme will have on the beliefs we form using that scheme. This could be called an 'epistemic' approach to conceptualization. Goodman (1955) argued that the value of a concept is given in the extent to which it allows us to form reliable expectations about its members, so that we should categorize in such a way that the resulting concepts best support inductive generalizations (see also Quine 1977). Psychologists have, similarly, proposed categorization metrics that would allow us to most reliably identify the kinds or features of objects we encounter (Rosch and Mervis 1975, Corter and Gluck 1992, Murphy 1982), or to formulate reliable theories about those objects (Chater 1999). And clustering algorithms in statistics and computer science have been developed to optimize the plausibility of analyses of bodies of data, again concerned with the epistemic outcomes of categorization (Thorndike 1953).

These latter approaches ultimately hinge on the idea that the value of a conceptual scheme can be understood in terms of the value of the beliefs we are inclined to form using that scheme, or, that concept utility can be best understood in terms of *epistemic utility*. We think this is the right way to approach the question. After all, it is often the case that approaches based on similarity or natural kinds are ultimately defended in epistemic terms, for example by the extent to which such schemes allow us to make plausible predictions. But how exactly should we think of the contribution a concept might make to the utility of a belief, and can we make this idea precise? The first thing we need to do is to decide what, in general, makes our beliefs valuable – something that has been the focus of the literature on *epistemic utility*. Let us turn to this literature next.

[2] Although sometimes these two can overlap, if natural kinds are themselves construed in terms of similarity, as in Homeostatic Property Cluster theory (Boyd, 1999).

2 Epistemic Utility

As Huber (2008) discusses, there are two distinct ways of thinking about the utility of beliefs. On one view, associated with Carnap (1962), a good belief or theory is one that is likely to be true. On this general approach, the measure of utility of a belief or theory is its plausibility.

Under a standard Bayesian analysis, the plausibility (p) of a hypothesis (H) is simply its probability given our evidence (E) and our other beliefs (B) about the world. Let us suppose, then, that p is a measure of the plausibility of any hypothesis, namely of its posterior probability conditional on the evidence and background beliefs:

$$p(H) := Pr(H|E \wedge B)$$

If we take p as an exhaustive measure of the value of any belief, then we should always prefer to adopt those hypotheses that score highest on this measure, and our task becomes that of figuring out how to evaluate p for different hypotheses.

However, as Popper (1959), Levi (1967), Maher (1993), and others have argued, plausibility cannot be the only thing we are concerned with when we decide what to believe. If it were, then we would have little explanation for the kinds of beliefs we are inclined to commit to. Rather than being solely concerned with plausibility, it would seem we are also concerned to acquire beliefs that are *informative* – that tell us something substantive about the world. Here is Maher on Cavendish's evaluation of his experiments on the weak electromagnetic force:

> "Consider the conclusion Cavendish drew from an experiment he conducted in 1773. The experiment was to determine how the electrostatic force between charged particles varies with the distance between the particles. Cavendish states his conclusion this way:
>
> *We may therefore conclude that the electric attraction and repulsion must be inversely [related] as some power of the distance between that of the* $2 + 1/50$*th and that of the* $2-1/50$*th, and there is no reason to think that it differs at all from the inverse duplicate ratio.* (Cavendish 1879, pp. 111-2).
>
> This statement indicates that Cavendish accepted Hc:
>
> (Hc) The electrostatic force falls off as the nth power of the distance, for some n between 1.98 and 2.02.
>
> Why wouldn't Cavendish have accepted only a weaker conclusion, for example by broadening the range of possible values of n, as in $H'c$:
>
> ($H'c$) The electrostatic force falls off as the nth power of the distance, for some n between 1.9 and 2.1.
>
> ...
>
> $H'c$ [is] more probable than the conclusion that Cavendish actually drew, as are infinitely many other weaker versions of Cavendish's hypothesis. The obvious suggestion

is that although these weaker hypotheses are more probable than Hc, they are also considerably less informative, and that is why Cavendish did not limit himself to these weaker hypotheses." (Maher 1993: 139-40)

Maher's point here is that if our goal in deciding what to believe were simply to maximize the chances of our having true beliefs, then we should water down our beliefs so that they were so weak as to be almost guaranteed to be true. If this were our only concern, indeed, then we should never adopt any beliefs other than tautologies, which are guaranteed to be true. Since we are concerned not just to have true beliefs but also informative ones, the utility of a belief should not be measured only in terms of plausibility, but also by its informativeness, or how much it tells us about the world.

How is informativeness measured? One way is to consider the number of possibilities a hypothesis rules out. Consider some hypotheses we might form about the outcome of throwing a die ten times over. The hypothesis that one toss will be an even number rules out just one alternative possibility – that all tosses will turn up an odd number – and is not very informative; the hypothesis that there will be a 5 and a 2 rules out more possible outcomes, and is more informative than the first; the hypothesis that three tosses will turn up a 6 rules out more possibilities again, and is again more informative. As we can see, the more possibilities a hypothesis rules out, the more informative it is.

The result is that informativeness can be measured in the same terms that we used to measure plausibility. As a hypothesis rules out more and more possibilities, after all, it becomes less and less plausible given the same evidence. The informativeness of a hypothesis will therefore vary inversely with its implausibility given our beliefs and evidence (for various other ways of thinking of informativeness see Huber 2008). Following Levi (1967), we may therefore adopt the following measure of informativeness:

$$i(H) := Pr(\neg H|E \wedge B)$$

Measuring informativeness in this way reflects Popper's (1959) idea that the utility of a belief is given by how 'falsifiable' it is, since the more possibilities a claim rules out, the more easily falsifiable it is.[3]

Of course informativeness alone cannot be what we are concerned with either, since false beliefs are not useful to us. It would seem that what we want, ideally, are beliefs that maximize both plausibility and informativeness (in principle such that $i(H) = p(H)$, but practically this may not be the case depending on which factor is viewed as more important in a given context). Following Huber (2008), we can call this the 'informativeness-plausibility' theory of acceptability.

Given that the informativeness of a belief varies inversely with its plausibility, these two 'virtues' of belief push in opposite directions. How do we decide what to believe,

[3]We refer to Huber (2008) for alternative and more fine-grained measures of informativeness. Levi's measure is subject to several objections, but suffices for our argument.

then – the more informative, or the more plausible of available hypotheses? A natural suggestion is that we accept the most informative hypothesis that meets our requirements on plausibility in a given context. For example, in a scientific context, we may have very little tolerance for error, and so we might restrict the claims we endorse only to those that meet a certain threshold of plausibility (such as the statistical thresholds that we are conventionally required to meet in scientific publications)[4]. Of those hypotheses that meet the threshold, we will endorse the most informative – rejecting other formulations of the claim that increase its strength but lower its plausibility (Maher 1993). Our tolerance for error can change, however. In a non-scientific context we may be ready to accept hypotheses that are much less plausible, but more informative. In a lay-context, for example, it might be much more useful to accept a broad generalization like 'birds fly and lay eggs', which is very informative, but far from strictly true.

There are many points of debate that could be explored further concerning epistemic utility, but that is not our purpose here (for further explorations see Levi 1967, Maher 1993, Huber 2008). Rather, allowing that we value both plausibility and informativeness in our beliefs, we wish to explore how we might evaluate the components of our beliefs – our concepts.

3 Concept Utility

Above we considered Cavendish's hypothesis about the electrostatic force. Let us consider another example of a scientific hypothesis – Rutz et al.'s (2016) hypothesis about Hawaiian Crows' tool-use abilities:

> Here we show that [...] the 'Alalā (*C. hawaiiensis; Hawaiian crow*), is a highly dexterous tool user. Although the 'Alalā became extinct in the wild in the early 2000s, and currently survives only in captivity, at least two lines of evidence suggest that tool use is part of the species' natural behavioural repertoire: juveniles develop functional tool use without training, or social input from adults; and proficient tool use is a species-wide capacity (Rutz et al. 2016: 403).

In this passage, Rutz et al. have committed to the following hypothesis H_{al}:

(H_{al}) Proficient tool use is a species-wide capacity in the 'Alalā.

This hypothesis, just like Cavendish's, could be weakened to make it more plausible. For example, given a high plausibility for H_{al}, we get an even higher plausibility for H'_{al}:

[4] Consider the Bayes Factor in Bayesian statistical analysis, of only accepting a hypothesis for which BF surpasses 10, or 30 (Jeffreys 1939). The Bayesian approach uses the same notion of plausibility we have adopted here, defined in terms of posterior probability of a hypothesis given evidence. In frequentist statistics, a different notion of plausibility is adopted, and conventional thresholds on p-values set similar boundary values regarding when to reject the null hypothesis.

(H'_{al}) Occasional tool use is a species-wide capacity in the ‘Alalā.

Since H_{al} entails H'_{al}, the latter is weaker than the former and therefore more probable given the same evidence. But of course, since H_{al} is stronger than H'_{al}, and already meets the standards of plausibility required in a scientific paper, we should endorse the former since it is more informative.

Notice, however, that there is another way of altering the informativeness and plausibility of the hypothesis – not by altering the strength of the claim made about the members of a particular class (the ‘Alalā), but by altering the range of the category the claim is extended to. That is, by altering the concept over which we project our inductive generalization. First, we can see that if we narrow the extension of the concept over which the generalization is projected, we weaken the hypothesis and increase its plausibility:

(H''_{al}) Proficient Tool use is a capacity to be found in the ‘Alalā that took part in our study.

H''_{al} is weaker than H_{al}, so it is more plausible given the same evidence. On the other hand it is less informative, since it tells us nothing about the ‘Alalā that did not take part in the experiment. Since we have no reason to think that ‘Alalā vary greatly in their cognitive abilities, H_{al} is supported by the evidence to a sufficiently high degree of probability to accept in a scientific context, and so Rutz et al have no reason to restrict their hypothesis to H''_{al}.

On the other hand, we could project our generalization over a concept with a broader extension:

(H'''_{al}) Proficient Tool use is a genus-wide capacity in Corvidae.

H'''_{al} is much stronger than H_{al}. The former entails the latter, and rules out many more possibilities – it rules out any question over whether Rooks (*Corvus frugilegus*) can use tools as well as the ‘Alalā , etc. It is clear why Rutz et al. do not propose H'''_{al}: our evidence about the ‘Alalā studied, coupled with our belief that all crow species might not be equally capable, makes H_{al} highly plausible, but not H'''_{al}, which extends the generalization to all crow species. So while this would be more informative, it would lower the plausibility to a level that we will not accept in a scientific study.

What this illustrates is that by changing the concept over which an inductive generalization is made, the informativeness and plausibility of the hypothesis changes. What exactly is it about the concept that co-varies with these changes?

First, the greater the range or extension of the concept, the greater the informativeness of the hypothesis. H'''_{al} is extremely informative, because it tells us about all sorts of different crows – Ravens, Rooks, Jackdaws etc. The first aspect of a concept that impacts on its epistemic utility will, then, be how many things the concept extends to – what we can call its *inclusiveness*. The more inclusive a concept, the more informative an inductive generalization made using that concept will be.[5] We define inclusiveness as the proportion of objects in a taxonomy that a concept extends to (see Appendix 1, Definition 2).[6] This gives us a first principle of concept utility:

Inclusiveness: The inclusiveness of a concept determines the informativeness of generalizations made using that concept.

What about plausibility? Clearly, in our example above, the plausibility of the generalization goes up as the range of things it is extended to narrows. But why is that? Falling under the concept 'ALALA, there is a smaller number of birds than fall under the concept CORVIDAE. But it isn't simply the cardinality of the category that has changed, it is the amount of variation that exists within the category. In the concept CORVIDAE there is a great deal of variation – if we discover something about 'Alalā , then we might doubt whether it will apply to Rooks, since we know that Rooks are different in many respects from 'Alalā. And while we decrease the variation within a category, the likelihood of discoveries about one object in the group extending to others increases. Since the members of the category 'ALALĀ are much more similar to one another than members of the broader category CORVIDAE are to one another, a discovery about one 'Alalā is more likely to apply to other 'Alalā than a discovery about one Corvid is to apply to other Corvids.

The second feature of a concept that affects the utility of generalizations it appears in is therefore what we might call its *homogeneity*, which will bear directly on the plausibility of beliefs we use the concept to form. We can define homogeneity in terms of the extent to which members of a category share features (see Appendix 1, Definitions 3 and 4). This gives us a second principle of concept utility:

[5] Note that our focus is on inductive or *ampliative* generalizations of the form 'all As are Bs' or 'many As are Bs', based on the observation of particular ABs and extending this to unobserved As. An existential generalization of the form 'some As are Bs', based on the observation of one or more ABs, is not inductive or ampliative, and for such an existential generalization, it is *not* the case that 'some As are Bs' is more informative than 'some Cs are Bs' when the concept A is more inclusive than C. Also, the inclusiveness and homogeneity of a concept only affect the informativeness and plausibility of such generalizations in this way when the concept occurs in the restrictor of the universal quantifier (i.e. in subject position): for example, 'all students smoke' is more informative than 'all blond students smoke', because 'student' is more inclusive than 'blond student' (a property described as the downward monotonicity of 'all' on its restrictor argument); on the other hand, 'all students smoke' is *less* informative than 'all students smoke cigars', even though 'smoke' is more inclusive than 'smoke cigars' – but in this case the more inclusive concept is in predicate position.

[6] Inclusiveness in this sense plays a role too in the categorization metrics of Rosch and Mervis (1975), and Corter and Gluck (1992), although the rationale is different from our own. For example Rosch (1978: 29) appeals to cognitive efficiency to motivate a preference for inclusive categories: since a conceptual scheme with highly inclusive categories will have fewer categories than one with less inclusive categories, a scheme with inclusive concepts will be easier to learn and use than the latter. Our concern here, on the other hand, is instead purely epistemic – even given a mind with unlimited cognitive resources, our reasons for appealing to inclusiveness would still stand, while practical reasons for appealing to inclusiveness might not.

Homogeneity: The homogeneity of a concept determines the plausibility of generalizations made using that concept.

Since we value both informativeness and plausibility in our beliefs, so too will we value both inclusiveness and homogeneity in our concepts. And just as informativeness and plausibility vary in inverse proportion to one another in beliefs, inclusiveness and homogeneity vary in inverse proportion in concepts (in the same way in which, classically, the 'extension' and 'comprehension' of a concept contravary, see Arnauld & Nicole 1662). This leads us to the following definition of concept utility (see Appendix 1, Definition 5):

Utility: The utility of a concept is the product of its homogeneity and inclusiveness.

As we now explore, maximizing concept utility as defined here can guide us in both the determination and revision of a conceptual scheme.

4 Determining and Revising a Conceptual Scheme

A conceptual scheme can be defined as an organization of objects into distinct categories. This organization could be into a single set of mutually exclusive categories which we can call a 'flat' organization. For example, we might be faced with a collection of objects that we could think of under the concepts ANIMAL, FURNITURE, and VEHICLE, where no concept is a sub-concept of any other. It could also be 'hierarchical' – we might be able to further subdivide the objects falling under ANIMAL between CAT and DOG, and divide the objects falling under FURNITURE between the concepts CHAIR and TABLE, etc. In this section and the next, we consider the problem of the determination and revision of 'flat' conceptual schemes, where concepts are mutually exclusive. We extend our treatment to hierarchical taxonomies in Appendix 2 of this paper.

4.1 Determining a Conceptual Scheme

Consider a domain consisting of three objects o_1-o_3. Suppose there are three relevant properties F_1-F_3 that are to be taken into consideration when we conceptualize these objects, which we can call 'features' (cf. Smith and Medin 1981, Corter and Gluck 1992).[7]

[7] Our model explains concept utility by assuming features to be given in the background, but it does not explain the selection of features, where related puzzles may arise concerning which features should be taken into consideration. We grant this is a further puzzle, but this is no reason to deny the importance of the notion of concept utility as we framed it. Given a selection of features, after all, one could still fail to carve out concepts in an optimal way in the absence of an understanding of concept utility.

	F_1	F_2	F_3
o_1	1	1	1
o_2	0	1	1
o_3	1	0	0

These objects could be anything at all. Various sea creatures, let us suppose, that display some salient features. Some have a spout-hole (F_1); some have a dorsal fin (F_2), and some have teeth (F_3). The question we are interested in is how the objects are to be clustered into distinct concepts, relative to that set of features. Is there an optimal way to partition the group?

We could conceptualize them as just one kind of thing, grouping all three objects under one concept (P_1). Or, we could think of them as three different kinds of thing – assigning a distinct concept to each object (P_3). Between those two extremes, there are three ways in which we could think of them as two kinds – grouping together o_1 and o_2 under one concept, and assigning o_3 to its own concept (P_{21}), or distinguishing o_1 from the others (P_{22}), or o_2 (P_{23}). In total, we have five possible conceptualizations or partitions of the domain (we use a vertical bar to delineate between cells):

$$\begin{array}{ll} P_1 & : \; o_1, o_2, o_3 \\ P_{21} & : \; o_1, o_2 \mid o_3 \\ P_{22} & : \; o_1 \mid o_2, o_3 \\ P_{23} & : \; o_1, o_3 \mid o_2 \\ P_3 & : \; o_1 \mid o_2 \mid o_3 \end{array}$$

Which one should we adopt? By calculating the utility of each partition in terms of our measure of concept utility, we shall see that one of the five emerges as optimal. To introduce our measurements, we will go through the calculation for one partition by step, and our reasoning should be easy to follow for subsequent cases.

Consider (P_{21}), which has two concepts, one of which includes the first two objects and the second of which includes the third:

$$P_{21}: \quad \{C_1 = \{o_1, o_2\}, C_2 = \{o_3\}\}$$

First let's consider the inclusiveness of the concepts in the partition – the proportion of objects in the domain that each concept extends to. C_1 includes two of the three objects in the domain, so it gets a value of inclusiveness of 2/3, $Incl(C_1)$=2/3. C_2 includes one of the three objects, and so $Incl(C_2)$=1/3.

Next we calculate the homogeneity of the concepts. We can think of this as the extent to which objects falling under the same concept are similar. This can be measured as the proportion of objects within a concept that possess a feature or lack it, whichever is bigger

– assuming that having a feature is just as much grounds for regarding things to be similar as lacking a feature.[8]

Let's see how this works by evaluating the homogeneity of the C_1 with respect to feature F_1. This feature is possessed by one of the objects in C_1 but the other lacks it. We want to say that they have no similarity with respect to F_1. And so we take the proportion of objects that have the feature, which is 1/2, and rescale it so that 1/2 becomes 0 (see footnote 7). This means that $Hom(C_1, F_1) = 0$. The second feature F_2 is shared by all objects in C_1, so $Hom(C_1, F_2)$=1. And that is the same for F_3. The homogeneity score for C_1 is the average of those three, so that $Hom(C_1) = (0+1+1)/3 = 2/3$. C_2 consists of only one object, so it is maximally homogeneous relative to each feature, so that $Hom(C_2) = 1$.

We can now combine the scores for inclusiveness and homogeneity to find a utility measure for each concept. For C_1 we find that: $U(C_1) = 2/3 \times 2/3 = 4/9$; and for C_2 we find that: $U(C_2) = 1 \times 1/3 = 1/3$. The average of the two gives us a score for this partition in terms of concept utility:

$$U(P_{21}) = 7/18$$

Consider for comparison the partition P_1 consisting of a single concept C encompassing all of the objects. Here the inclusiveness will be 1, $Incl(C) = 1$. Each feature is shared by two thirds of the objects, giving each feature a homogeneity score of $Hom(C, F_i) = 1/3$ by our scaling algorithm, so that $Hom(C) = 1/3$. From this it follows that $U(P_1) = U(C) = 1/3$. Grouping all three objects together therefore gets a slightly lower score than splitting them in two. The reason is that although the inclusiveness of the single concept in P_1 is 1, the homogeneity is just 1/3, because this single concept now groups together all three objects, one of which has little in common with the other two. Although P_{21} scores much lower on inclusiveness by splitting the domain into two concepts, the gain in homogeneity that the division into two concepts results in gives it a higher overall score. In fact, P_{21} beats all of the other partitions (see Appendix 2 for calculations):

$$U(P_{22}) < U(P_{23}) < U(P_1) = U(P_3) < U(P_{21})$$

These results make intuitive sense. P_{22} scores the lowest, because it groups together two objects that have no feature in common. P_{23} does slightly better by grouping two objects that have one feature in common, while P_{21} scores the highest by grouping together the two objects that have most in common. P_1 and P_3 are in a tie because they trade off inclusiveness for homogeneity and conversely – P_1 gets the highest score for inclusiveness

[8]By measuring homogeneity in terms of having or lacking a common feature, the resulting overall homogeneity score will always be a 1/2 or more, since it is not possible for less than half the objects in any group to either lack or possess some feature (if 0.1 of the group lacks the feature, then 0.9 possess it, hence the proportion of a group that possess or lack a feature can never be less than 0.5). However, the minimal value on a scale should ideally be represented as zero. For this reason we rescale the homogeneity values so that 1/2 is represented as 0, 1 is represented as 1, and other values fall in between, giving us a more sensitive measure. This is done by multiplying the value by 2 and subtracting 1 (see Appendix 1, Def. 3).

but the lowest score for homogeneity by including all objects in a single concept, while P_3 gets a maximal score for homogeneity but gets the lowest score for inclusiveness, by partitioning the domain into three concepts.

This shows that when faced with multiple ways of categorizing the objects in a domain, measuring the inclusiveness and homogeneity of different classifications gives us a principled way to choose between them. Now let's see how maximizing the combination of those values translates back into maximizing the utility of the beliefs we use these concepts to form. Suppose for a moment that the objects o_1, o_2, o_3 each stand for populations of, let's say, 100 objects bearing those features. Now suppose that you make a new discovery about one member of the group denoted as o_3 – you notice that it has a pentadactyl bone structure in its fins. You are now inclined to expect that other creatures might have pentadactyl limbs given that you have observed one with this feature. Over which individuals do you project this generalization?

This will depend on which conceptualization you have adopted, assuming you will generalize the discovery to the category to which you have assigned o_2. In the case of P_{22}, you would generalize over other individuals denoted by o_2 and also those denoted by o_3. This is a relatively informative inference, telling you about 200 creatures. But because the creatures denoted by o_2 and those denoted by o_3 are very dissimilar, failing to share any features considered so far, we should not expect this inference to be very plausible. Consider a similar generalization made in P_{21}, where the objects denoted by o_2 were grouped together with those denoted by o_1. If we extend the generalization to all members of this category, it will be just as informative as before, again telling us about 200 creatures. But it will also be much more plausible, because the members of this category are much more similar. And so we can see how optimizing concept utility in turn optimizes epistemic utility, conceived in terms of both plausibility and informativeness.

4.2 Revising a Conceptual Scheme

So much for the determination of a conceptual scheme. We now consider two ways in which discoveries about one's environment can justify the revision of a conceptual scheme. First, the discovery of new *features* in an environment can justify such a revision. This shouldn't be surprising – given closer examination of objects in our environment, it is not unusual to find out that objects that appeared closely related at a glance are actually quite different, or that objects that appeared very different initially turn out to have more in common than we realized. And such considerations may prompt us to revise our concepts, as Waismann (1945) suspected when he proposed that our concepts need to have an 'open texture', to accommodate the discovery of new features in an environment.

To illustrate how new features can motivate a revision on our current account, consider what happens if we add two features, F_4 and F_5, to the previous matrix:

	F_1	F_2	F_3	F_4	F_5
o_1	1	1	1	1	1
o_2	0	1	1	0	0
o_3	1	0	0	1	1

Let's imagine these newly observed features in our population of sea creatures are limb bone structure (F_4) – two a horizontal tail but the other doesn't; and feeding habits (F_5) – two eat krill, the other doesn't. While bearing in mind just the original three features, we found that o_2 and o_1 had more in common than either had with o_3; but now with these further features in mind it turns out that o_1 and o_3 have more in common than either has with o_2 (o_1 and o_3 are supposed to be whales, while o_2 is a shark). This has a clear impact on the optimality of the competing conceptual schemes. While before P_{21} scored highest, now the highest score is attained by P_{23}, which groups together objects o_1 and o_3, as the reader can check for herself. The result is that taking maximal concept utility as a goal for conceptualization provides us with a standard that can motivate the revision of a conceptual scheme in light of the discovery of new features.

Importantly, the discovery of new *objects* in a domain, without discovering any new features, can also justify such revisions. Consider another object-feature matrix:

	F_1	F_2	F_3
o_1	1	1	1
o_2	1	1	1
o_3	1	1	1
o_4	0	0	1

Here we have found ourselves in an environment with three objects that are identical with respect to the features F_1-F_3, and a fourth that differs from the others with respect to the first two features, but is similar with respect to the third. First let us decide how we should conceptualize the group - as falling under one single concept, or perhaps splitting the group so that o_4 is distinguished from the others. Consider a simple partition P_1 that contains just one concept $C = \{o_1, o_2, o_3, o_4\}$. The inclusiveness of the single concept in this partition is 1, and its homogeneity is 2/3, so its utility $U(P_1) = 2/3$. Now consider a partition P_2 that groups the first three objects together under one concept $C_1 = \{o_1, o_2, o_3\}$, and assigns the fourth to a separate category $C_2 = \{o_4\}$. C_1 includes 3/4 of the objects and has a homogeneity of 1, while C_2 scores 1/4 for inclusiveness and 1 for homogeneity, so the utility of the partition is is $U(P_2) = 1/2$. This is less than $U(P_1) = 2/3$, and so here it is optimal to think of the objects as just one kind of thing. In this example, then, a domain includes an 'oddball' that differs from the other objects in the domain, but its difference from the others is not sufficient to justify splitting the domain into two concepts. Precisely because we value inclusiveness in addition to homogeneity, we prefer not to split in this case, even though doing so would increase the overall homogeneity.

But consider what happens if we expand the domain by including more objects similar to the oddball, without adding any new features:

	F_1	F_2	F_3
o_1	1	1	1
o_2	1	1	1
o_3	1	1	1
o_4	0	0	1
o_5	0	0	1
o_6	0	0	1

The utility of a single partition that groups all the objects together has now dropped to $U(P_1) = 1/3$, even though no new features have been added: $Inc(P_1)$=1, but $Hom(P_1) =$ 0+0+1. The reason for this is that in the original domain, when all the objects are grouped together, three quarters of the objects in that concept share all their properties. But given the increase in the number of objects similar to the oddball, fully half of the objects are now distinct from the others with respect to two thirds of the properties, F_1-F_2. A partition that was originally quite homogeneous can therefore lose its homogeneity without any new features appearing among its members, but simply because new objects are added to the domain that have the same features as already existing objects. The utility of a partition that distinguishes the first three objects from the others remains, however, at $U(P_2) = 1/2$: $Inc(P_2)$=1/2, but $Hom(P_2) = 1$. This is now a higher score than the utility of a single concept partition, so that it has become optimal to split the domain.

Changes in the *proportion* of objects with particular features in a domain can in this way justify revising a conceptual scheme, without any new features being added. Next, we turn to the puzzling revision of the concept PLANET, and argue that its revision follows exactly the pattern just described.

5 The Case of Planet

In 2006, the International Astronomical Union formed a committee to resolve a growing dispute concerning the category PLANET. During the convention of the IAU, two resolutions were submitted to a vote and adopted, Resolutions B5 and B6. The effect of these resolutions was to alter the definition of the category so that Pluto and several other newly discovered celestial objects were no longer to count as planets.

In the opinion of some philosophers, this dispute was merely terminological. Chalmers (2011), for example, argues that it was essentially a *verbal dispute* – a question of language, rather than a question of fact. Among astronomers, on the other hand, the case is thought of quite differently. Mike Brown, one of the astronomers centrally involved in the discoveries that led to the demotion of Pluto wrote: ‘the debate about whether or not Pluto is a planet is critical to our understanding of the solar system. It is not semantics. It is fundamental classification' (Brown 2010: 232).[9] In other words, Brown regarded the dispute as driven by truth – not terminological convenience. Our view is that Brown was right – that the inclusion of Pluto in the category PLANET turned out to be factually incorrect, assuming that the goal of such a categorization is to maximize epistemic utility. Our account of concept utility can now be recruited to explain why.

5.1 Context of the IAU Resolutions: The Stern-Levison Criteria

Let us recall a few facts concerning the background to the 2006 decision. First, the concept PLANET had undergone some significant changes before 2006. In Ancient astronomy, the concept extended to apparently moving or 'wandering' celestial bodies, and therefore included both the Sun and the Moon, but not the Earth. With the switch to a Heliocentric system, PLANET came to include the Earth, but no longer the Sun, nor the Moon. That change depended on the introduction of a new criterion for counting as a planet, namely orbiting the Sun.

During the nineteenth century, new celestial bodies were discovered, and some of these were first regarded as planets, such as Ceres, discovered in 1801. After the discovery of Pallas in 1802, however, a celestial body of roughly the dimensions of Ceres, the concept ASTEROID was proposed by Herschel in 1802. Ceres was reassigned to this new category along with Pallas. This decision was based on the observation that both had significantly smaller sizes than the other planets, and therefore seemed to form a distinct class of entities (Soter 2006). As we shall see the revision of the concept that took place in 2006 followed a similar pattern.

In 2005, Brown announced the discovery of a new body in the solar system, 2003-UB313. This object, which came to be called 'Eris', appeared to qualify as a planet under the definition at the time, in particular because it was slightly larger than Pluto. It quickly became clear to Brown and several other astronomers, however, that calling Eris a planet could mean that a potentially large number of other celestial bodies should also be included under the concept, which would undermine the overall homogeneity of the category (Brown 2006). These included a number of celestial bodies discovered a few years before Eris, in the region of the Solar System called the Kuiper Belt. Brown began to suspect that neither Eris nor Pluto should not be thought of as a planet after all.

In fact, astronomers had long recognized the distinctness of Pluto from the other 8 planets. As early as 1930, astronomers had pointed out the significantly tilted orbit of Pluto, of about 17° relative to the other 8 planets; the fact that the orbit of Pluto crosses that of Neptune; the much greater distance between Pluto and Neptune relative to the distances between the other planets; and the unusually icy character of Pluto (Leonard 1930, Brown 2010). In 2000, the astronomers Stern and Levison identified two more features

[9] Brown did not take part in the IAU vote, but approved of the outcome, see Brown (2006), and Brown (2010).

typical of planets, one of which set Pluto apart even further. The first is what they called a *physical criterion* – for a body to have sufficient mass to reach hydrostatic equilibrium. This is reached when an object is sufficiently massive that its own gravity causes it to take on the shape of a sphere. This criterion is inclusive of Ceres, Pluto and some other Kuiper Belt Objects discovered at the time. The second is what they call a *dynamical criterion*, which is satisfied by 'a body in orbit about a star that is dynamically important enough to have cleared its neighboring planetesimals in a Hubble time' (Stern and Levison 2002: 4). To satisfy the dynamical criterion, a body must have cleared other objects such as rocks or debris from its orbital path. The objects in the Kuiper belt will not qualify, because none of these have sufficient gravity to pull the others into its own mass. Pluto did not satisfy this criterion, but the 'traditional' 8 planets did. [10]

Stern and Levison's criteria came to play an important role in the 2006 IAU resolutions.[11] The first resolution, B5, defines a planet 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 neighbourhood around its orbit."

Criterion (a) is the criterion underlying the traditional heliocentric scheme. Criterion (b) corresponds to Stern and Levison's physical criterion, and criterion (c) corresponds to their dynamical criterion. On the basis of that definition, which has the effect of excluding Pluto from the category PLANET, Resolution B6 posits a new category of celestial objects – TRANS-NEPTUNIAN OBJECT – of which Pluto is presented as the prototype.[12]

The difference this made can be seen in our simplified versions of the taxonomies endorsed by the IAU in 1930 (Figure 1) and after the 2006 resolutions (Figure 2). In 1930, a threefold distinction is made between the planets, including the 'Gas Giants' Jupiter, Neptune, Uranus and Saturn; the 'Terrestrials' Mercury, Venus, Earth and Mars; and Pluto, which had sometimes been characterised as an 'Icy Dwarf'. By 2006, the 'Icy Dwarf' cat-

[10]Brown proposed a similar criterion for counting as a planet, based on the relation between the mass of a celestial body and the total mass of the bodies orbiting around it; and shortly before the IAU resolutions were voted, Soter submitted a paper in which he proposed a synthesis of both ideas in the form of a 'planetary discriminant', defined as the ratio of the body mass to the aggregate mass of the neighboring bodies in its orbit (see Soter 2007 for an overview). We leave this detail to one side because Soter's discriminant is very well correlated with Stern and Levison's discriminant, motivating more or less the same category division.

[11]See `https://www.iau.org/static/resolutions/Resolution_GA26-5-6.pdf`.

[12]Oddly, in Resolution B6 Pluto is also assigned to a further novel category of DWARF PLANET which passes criteria a) and b) but not c). However, the IAU Resolution B5 implies that DWARF PLANET, despite the name, is in fact a separate category rather than a subcategory of PLANET, so that using the term DWARF PLANET is misleading (see Brown 2010 for a criticism of that inconsistency).

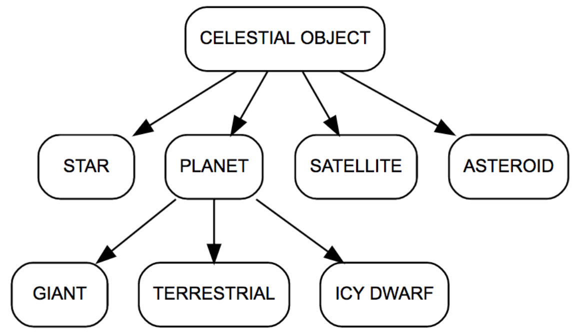


Figure 1: 1930 taxonomy for celestial bodies

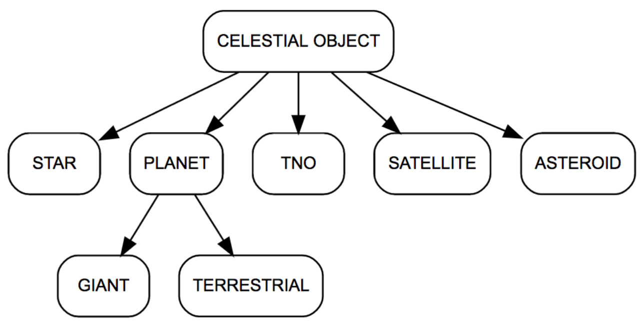


Figure 2: 2006 taxonomy for celestial bodies

egory has been eliminated, and Pluto and its companions have been moved into the new category TRANS-NEPTUNIAN OBJECT (TNO).

What is striking about the pattern of decisions is that the features that are now used to distinguish TRANS-NEPTUNIAN OBJECT from PLANET were already known to distinguish Pluto from the other planets in 2000 – but at that time, they did not seem to provide sufficient reason to exclude Pluto from the category. [13] It was only once new objects that were similar to Pluto were discovered did the tide of opinion among astronomers began to change, and the criteria that before had seemed insufficient to exclude Pluto from the category now seemed to justify this move. If homogeneity or similarity alone was what we were concerned with in devising a conceptual scheme, however, Pluto should have been

[13] Stern and Levison had in fact proposed that Pluto might be thought of as belonging to a subcategory UNTERPLANET, while those objects that satisfy their dynamical criterion might be thought of as ÜBERPLANETS.

Celestial Body	$\Lambda > 1$
Mercury	1
Venus	1
Earth	1
Mars	1
Jupiter	1
Saturn	1
Uranus	1
Neptune	1
Pluto (1930)	0
Quaoar (2002)	0
Sedna (2003)	0
Eris (2003)	0
Orcus (2004)	0
Makemake (2005)	0

Table 1: Satisfaction of Stern and Levison's criterion Λ

excluded from the category PLANET long before the discovery of the other TNOs, when its distinctiveness was already understood. How do we make sense of this?

5.2 Rationalizing the Revision

In Table 1, the first nine bodies you will see are the 9 planets according to the taxonomy of planets in 2000, which had been in place since the discovery of Pluto (1930). Underneath these are 5 new celestial bodies discovered by Brown and his team between 2000 and 2005, including Eris. Before the discovery of the five lower Kuiper Belt objects in the table, the line demarcating the category PLANET falls below Pluto. After the discovery of those objects, it is moved above Pluto. Each planet is assigned a 1 or a 0 depending on whether it satisfies the Stern-Levison dynamical criterion Λ.

Our explanatory challenge is to show why it was that before the discovery of the new objects, excluding Pluto from the category PLANET was not justified; but that once these objects were discovered, the move becomes justified – and that on the basis of features that were already known to distinguish Pluto from the other planets at the earlier time.

First let's consider ways to categorize the domain as it was known in 2000, which includes just the first 9 objects. One way would be to split the group in two – one concept for Pluto, and another for the remaining 8 planets. Relative to the domain, the inclusiveness of a concept extending only to Pluto would be 1/9 (since it contains only 1 of the 9 objects in the domain) and its homogeneity 1 (since it is perfectly homogeneous, having only one member). Its utility is therefore 1/9. The utility of a category for the other 8 bodies is

8/9 for symmetric reasons (and abstracting away from other ways in which the remaining 8 planets differ, for the sake of argument). The overall utility of a partition into these two concepts is therefore $(1/9 + 8/9)/2 = 1/2$. By contrast, consider a partition including just one concept that covers all nine objects. Its inclusiveness is 1, since it now includes all the objects under consideration. Its homogeneity relative to Λ is 8/9, since 8 out of the 9 objects share the discriminant property, which scaled according to our algorithm is $2 \cdot (8/9) - 1 = 7/9$. When the known domain consisted of only the first nine objects, then, it was optimal not to separate Pluto from the others, and for exactly the same reason considered above (section 4.2): because inclusiveness matters in addition to homogeneity, a large loss in inclusiveness (reducing its value from 1 to 1/2 by splitting) is not worth a small gain in homogeneity (increasing its value from 7/9 to 1), and sometimes it's better to retain an oddball member of a category to maximize overall utility.

Consider now the expanded domain five years later in 2005, given the discovery of Eris and the other four Kuiper Belt objects. Once again we can consider the utility of a partition into two concepts: one equivalent to the old concept PLANET encompassing the first nine objects including Pluto, which we can call O (for old), and a separate concept for the new objects, let's call this N (for new). The inclusiveness of O is 9/14, and its homogeneity again is 7/9. The inclusiveness of N is 5/14, and its homogeneity is 1 (they all lack the discriminant). From our definitions, it follows that the utility of O is $9/14 \cdot 7/9 = 1/2$ whereas the utility of N is 5/14. The overall utility of that partition is therefore $(1/2 + 5/14)/2 = 6/14$, or 3/7.

On the other hand, consider a partition of this newly expanded domain including a concept TNO that groups Pluto together with the new objects, and a concept PL that includes just the first eight bodies. For this partition, $Incl(PL) = 8/14$, and $Hom(PL) = 1$; $Incl(TNO) = 6/14$, and $Hom(TNO) = 1$ – with respect to the second Stern-Levison criterion, both groups are now perfectly homogeneous. The utility of PL is now 8/14, and the utility of TNO is 6/14, hence the overall utility of that partition is 1/2. This beats the overall utility of the division that keeps Pluto with the planets, which is 3/7. In other words, before the new objects are discovered, assigning Pluto to a distinct category from the traditional 8 planets has a lower utility than keeping it in a single category with them; but after the new objects are discovered, separating Pluto from the other planets and assigning it to the TNO category has a higher utility.

Note finally that both solutions are better than 'stretching' the category PLANET to encompass all 14 celestial bodies under a single concept. For even though the inclusiveness of the corresponding concept would be 1, its homogeneity would fall to $2 \cdot (8/14) - 1 = 1/7$, so the overall epistemic utility of that scheme would itself be 1/7, as a result of the category becoming too heterogeneous.[14]

The case of PLANET therefore follows the pattern discussed at the end of section (4.2),

[14]Supporting Brown's intuition that a 'leave no iceball behind' option of including all potential candidates for planethood would ultimately create an unweildy category (Brown 2006).

where a concept includes an 'oddball' member, but the resulting heterogeneity is not so severe as to justify splitting the concept. Because we value inclusiveness in addition to homogeneity, a small gain in homogeneity need not justify splitting a category. Once sufficiently many objects similar to the oddball member are discovered, however, the cost to inclusiveness of splitting the category can be offset by a now greater gain in homogeneity. Understood in these terms it also becomes clear that we have here no merely terminological dispute (cf. Chalmers 2011): the distribution of properties and objects in the domain under consideration means that some ways of conceptualizing the domain are measurably better than others. Including Pluto in the category PLANET would fail to provide us with an optimal conceptual scheme, as a matter of fact.

Admittedly, our analysis simplifies the complexity of the original case, since many more features vary across the planets and TNOs than just the Stern-Levison criterion. Nevertheless, it is primarily on the basis of this criterion that the IAU ultimately came to exclude Pluto from the category PLANET. Our analysis also considerably shrinks the domain of relevant objects, since by 2000 dozens of so-called Kuiper Belt objects had already been discovered. If we trust Brown's testimony, however, it is indeed the discovery of those first 'large' Kuiper belt objects between 2000 and 2005 that gradually put pressure on the old conceptual scheme, and led to its revision. And so we think that we have identified the crucial elements of the transition, and the factors that really lead to the revision of the conceptual scheme.

6 Conclusion

Understanding the utility of a conceptual scheme is surely of central concern to the theory of concepts, to accounts of scientific classification, and indeed to any area in which categorization plays an important role. In spite of this it has seen relatively little philosophical discussion, perhaps because it has been unclear how to investigate the question. We have shown here how a measure of concept utility can be derived from broader notions of epistemic utility, resulting in an account can be applied to any domain of categorization.

Our account departs from extant proposals in several ways. First, it does not lean primarily on similarity to decide on category boundaries, but recognizes that similarity, here cashed out in terms of homogeneity, is only part of what we are concerned with in a conceptual scheme. Second, it does not appeal to naturalness as a competing consideration, thereby avoiding difficulties concerning the multiplicity of natural distinctions. Focusing instead on the homogeneity and inclusiveness of concepts, our account recognizes the importance of both plausibility and informativeness to the beliefs that a conceptual scheme will have an impact on, and thereby provides a more robust and predictive account of concept utility.

Several issues remain to be investigated. The psychological accounts of concept utility discussed at the outset have been designed to explain human categorization tendencies, in particular our inclination to always categorize an environment at a predictable level of fineness of grain, sometimes called the ‘basic level’ (see Murphy 2002 chapter 4 for discussion). We expect that our account will be equally predictive of performance in such tasks. Indeed, since these accounts are largely focused on maximizing the homogeneity of conceptual schemes, their predictions will differ in important ways from our own. By recognizing the importance of inclusiveness in addition to homogeneity, our model predicts that the proportion of members of objects that share properties in a domain will affect the way we are inclined to categorize, something that accounts primarily based on homogeneity will not predict. Whether these predictions will hold up in experimental categorization tasks is a question we leave for future investigation.

References

Arnauld, A. & Nicole, P. (1662). *Logic or the Art of Thinking.* J. V. Buroker (ed.). Cambridge University Press.

Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. Wilson (ed.) *Species: New Interdisciplinary Essays.* Cambridge: MIT Press: 141-185.

Brown, M. (2006). The Eight Planets. `http://web.gps.caltech.edu/~mbrown/eightplanets/`.

Brown, M. (2010). *How I killed Pluto and Why it Had it Coming.* Random House.

Carnap, R. (1928). *Der Logische Aufbau der Welt*, Leipzig: Felix Meiner Verlag.

Carnap, R. (1962). *Logical Foundations of Probability.* 2nd ed. Chicago: University of Chicago Press.

Cavendish, H. (1879). *The Electrical Researches of the Honourable Henry Cavendish*, ed. by J. Clerk Maxwell. Cambridge.

Chalmers, D. (2011). Verbal Disputes. *Philosophical Review*, 120: 515-66.

Chater, N. (1999). The Search for Simplicity. *The Quarterly Journal of Experimental Philosophy* 52 A(2): 273-302.

Corter J. E. & Gluck M. A. (1992). Explaining basic categories: feature predictibility and information. *Psychological Bulletin* 111 (2): 291–303.

Cracraft, J. (1983). Species Concepts and Speciation Analysis. In *Current Ornithology*, R. Johnston (ed.), New York: Plenum Press: 159–187.

Dupré, J. (1993). *The Disorder of Things : Metaphysical Foundations of the Disunity of Science*, Cambridge MA: Harvard University Press.

Ereshefsky, M. (1998). Species Pluralism and Anti-Realism. *Philosophy of Science*, 65: 103–120.

Frege, G. (1891). Funktion und Begriff. Reissued in G. Frege, *Funktion, Begriff, Bedeutung.* Kleine Vandenhoeck Reihe, 1980.

Gärdenfors, P. (2000). *Conceptual Spaces.* MIT Press.

Goodman, N. (1955). *Fact, Fiction and Forecast.* Harvard University Press: Cambridge MA.

Gopnik, A., & Meltzoff, A. (1997). *Words, Thoughts, and Theories*, Cambridge, MA: MIT Press.

Huber, F. (2008). The Plausibility-Informativeness Theory. In V.F. Hendricks & D.H. Pritchard (eds.), *New Waves in Epistemology.* Basingstoke: Palgrave Macmillan: 164-191.

International Astronomical Association. (2006). Resolutions B5 and B6 adopted at the general assembly. `https://www.iau.org/static/resolutions/Resolution_GA26-5-6.pdf`.

Jeffreys, H. (1939). *Theory of Probability.* Oxford: Clarendon Press.

Keil, F. (1989). *Concepts, Kinds, and Cognitive Development.* Cambridge, MA: MIT Press.

Kripke, S. (1980). *Naming and Necessity.* Harvard University Press.

Leonard, F. C. (1930). The New Planet Pluto. *Astronomical Society of the Pacific Leaflets*, 1 (30): 121-124.

Levi, I. (1967), *Gambling With Truth. An Essay on Induction and the Aims of Science.* London: Routledge.

Maher, P. (1993). *Betting on Theories.* Cambridge University Press.

Mayr, E., (1969). *Principles of Systematic Zoology.* New York: McGraw-Hill.

Murphy, G. L. (1982). Cue validity and levels of categorization. *Psychological Bulletin*, 91, 174-177.

Murphy, G. (2002). *The Big Book of Concepts.* MIT Press.

Popper, K. (1959). *The Logic of Scientific Discovery.* Routledge.

Putnam, H. (1975). The Meaning of Meaning. *Minnesota Studies in the Philosophy of Science* (7) 131-193.

Rosch, E., & Mervis, C. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, 7: 573-605.

Rosch, E. (1978). Principles of Categorization. In Rosch, E. and Lloyd, B. (eds.) *Cognition and Categorization*, Laurence Erlbaum.

Rutz, C., Klump, B. C., Komarczyk, L., Leighton, R., Kramer, J., Wischnewski, S., Sugasawa, S., Morrissey, M. B., James, R., St Clair, J., Switzer, R. A. & Masuda, B. M. (2016). Discovery of Species wide tool use in the Hawaiian Crow. *Nature.* 537, 7620, p. 403-407.

Smith, D. and Medin, S. (1981). *Categories and concepts.* Harvard University Press.

Soter, S. (2006). What is a Planet? *The Astronomical Journal.* 132 (6): 2513-2519.

Soter, S. (2007). What is a Planet? *Scientific American*, January 2007: 34-41.

Stern, S. Levison, H. F. (2002). Regarding the criteria for planethood and proposed planetary classification schemes. *Highlights of Astronomy* 12: 205–213. (Presented at the IAU Assembly 2000).

Thorndike, R. L. (1953). Who belongs in the family?. *Psychometrika* 18 (4): 267-276.

Waismann F. (1945). Verifiability. *Proceedings of the Aristotelian Society* 19: 119-150.

Appendix 1: Basic definitions

Following Frege, we may view concepts as functions from objects to truth values. For simplicity, however, we identify them with their extensions, hence with the *categories* they determine.[15] We define concepts relative to objects and to features (see Smith and Medin 1981, Corter and Gluck 1992). Features are unary properties. They can be viewed as concepts therefore, but concepts in our approach generally correspond to combinations of features, so features are intended to be more primitive in that sense.[16] For simplicity, we always assume the objects and the features to be finite.

Definition 1. *A concept C is a function from objects of a domain to truth values. For each domain, it determines a category, that is a subset of the domain D (i.e. the set of C-objects, or objects satisfying C). For simplicity we identify concepts with the categories they determine in what follows.*

Definition 2. *The inclusiveness of a concept C, noted $Incl(C)$, is the proportion of objects of D satisfying C.*

Definition 3. *The homogeneity of a concept C relative to feature F_i, written $Hom(C, F_i)$ is the proportion of the C-objects positively satisfying feature F_i, or the proportion of C-objects not satisfying feature F_i, whichever is greater, rescaled to a minimum value of 0 and a maximum value of 1 as follows: when the higher proportion is x, the homogeneity is $2x - 1$.*

Definition 4. *The homogeneity of a concept C relative to a finite set of features $(F_i)_{i \leq n}$ (written $Hom(C)$ when feature set is clear from context) is the sum of the homogeneities of C relative to each feature, divided by the number n of features.*

Definition 5. *The epistemic utility of a concept relative to a set of features is the product of its inclusiveness and homogeneity relative to that set, namely:*

$$U(C) = Incl(C) \times Hom(C)$$

Definition 6. *A partition is a set of nonempty subsets of D that are mutually exclusive and exhaustive of D. Given a partition P of D into distinct concepts, the epistemic utility of P is the average of the utilities of the concepts in P.*

Appendix 2: Hierarchical taxonomies

We know how to calculate the epistemic utility of a simple partition, but what about more complex taxonomies, involving a hierarchy of levels? We may define a taxonomy as follows:

[15]See Frege (1891) on concepts as functions.
[16]On features as primitives of a theory of complex concepts, see Smith and Medin (1981).

Definition 7. *A taxonomy is a finite family* $(P_i)_{i \leq m}$ *of partitions of the domain into distinct concepts (categories), such that for each* i*, the partition* P_{i+1} *is a refinement of* P_i *(a partition refines another one if every concept of the first is a subset of a concept in the second). The level* k *in a taxonomy is the corresponding partition* P_k*.*

We can construct an illustration from the 'flat' partitions considered in section 3. There we considered a domain consisting of three objects relative to three properties:

	SPOUT	FIN	TEETH
o_1	1	1	1
o_2	0	1	1
o_3	1	0	0

We noted that there are five possible partitions of this domain:

$$\begin{array}{lcl} P_1 & : & o_1, o_2, o_3 \\ P_{21} & : & o_1, o_2 \mid o_3 \\ P_{22} & : & o_1 \mid o_2, o_3 \\ P_{23} & : & o_1, o_3 \mid o_2 \\ P_3 & : & o_1 \mid o_2 \mid o_3 \end{array}$$

These same partitions can now be 'stacked' to form three different taxonomies, T_1-T_3. The highest level in each is equivalent to P_1, and the lowest to P_3, while there are three different ways to conceptualize the objects at the second level. L_2 in T_1 is equivalent to P_{21}, L_2 in T_2 is equivalent to P_{22} and L_2 in T_3 is equivalent to P_{23}:

$$T1 \left\{ \begin{array}{c|ccccc} L_1 & o_1 & & o_2 & & o_3 \\ \hline L_2 & o_1 & & o_2 & \mid & o_3 \\ \hline L_3 & o_1 & \mid & o_2 & \mid & o_3 \end{array} \right.$$

$$T2 \left\{ \begin{array}{c|ccccc} L_1 & o_1 & & o_2 & & o_3 \\ \hline L_2 & o_1 & \mid & o_2 & & o_3 \\ \hline L_3 & o_1 & \mid & o_2 & \mid & o_3 \end{array} \right.$$

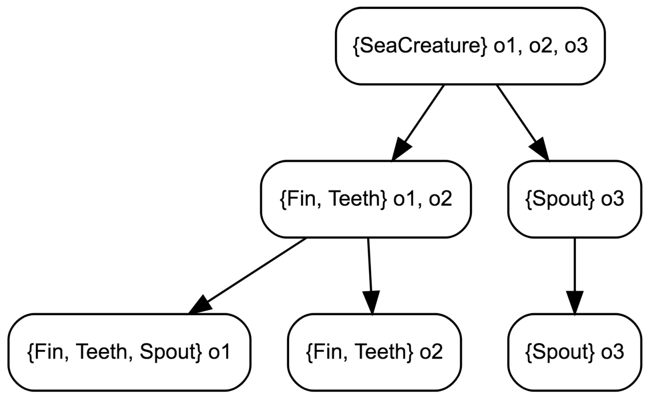
$$T3 \left\{ \begin{array}{c|ccccc} L_1 & o_1 & & o_3 & & o_2 \\ \hline L_2 & o_1 & & o_3 & \mid & o_2 \\ \hline L_3 & o_1 & \mid & o_3 & \mid & o_2 \end{array} \right.$$

Figures 3 to 5 give a more explicit represent of T_1 to T_3 both in tree form and in the form of a dataframe.[17] In each taxonomy there are three *levels*, and six *concepts*, based on three binary *features* (Fin, Spout, Teeth). Each node corresponds to a concept, the objects included are listed, and within curly brackets are the features that all objects in that concept have in common, which in some cases is none, as in T_2 (Figure 4) (an exception is the root node, which we label with the feature {SeaCreature} for clarity, although we do not include this general feature in our calculations).

Below the trees, a table illustrates an implementation in R of our algorithm for computing utility. Each row corresponds to a node or concept of the taxonomy, and in columns are marked its level, its cardinality, and its homogeneity relative to each relevant feature of the taxonomy. The end columns Hom, $Incl$, and U give the overall homogeneity, the inclusiveness and the epistemic utility of each node.

No unique method exists to assign epistemic utility to a taxonomy, but one of the simplest is to calculate the epistemic utilities of all its levels, where the utility of the level is calculated as before for partitions, and average these. On that method, since T_1, T_2 and T_3 differ only at level-2, it follows from our previous calculations that T_1 has higher utility than T_3, and T_3 higher utility than T_2. These values are indicated in a separate table underneath each dataframe.

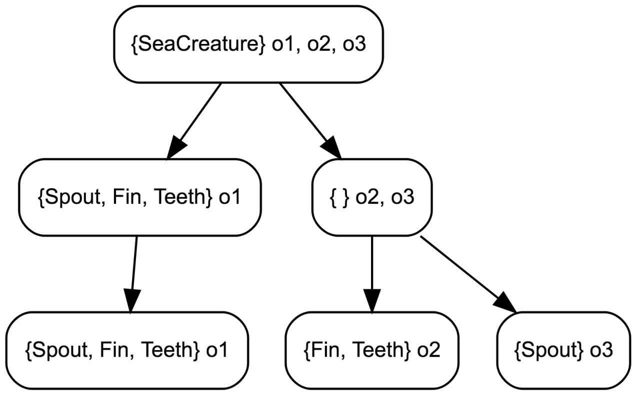
[17] We derived both by implementing the algorithm presented in Appendix 1 in R using the library data.tree (created by Christoph Glur).



	T1	level	card	n_SPOUT	n_FIN	n_TEETH	Hom	$Incl$	U
1	{SeaCreature} o1, o2, o3	1	3	2.00	2.00	2.00	0.33	1.00	0.33
2	|—{Fin, Teeth} o1, o2	2	2	1.00	2.00	2.00	0.67	0.67	0.44
3	| |—{Fin, Teeth, Spout} o1	3	1	1.00	1.00	1.00	1.00	0.33	0.33
4	| °—{Fin, Teeth} o2	3	1	0.00	1.00	1.00	1.00	0.33	0.33
5	°—{Spout} o3	2	1	1.00	0.00	0.00	1.00	0.33	0.33
6	°—{Spout} o3	3	1	1.00	0.00	0.00	1.00	0.33	0.33

level	1	2	3	T_1
U	0.333	0.388	0.333	0.351

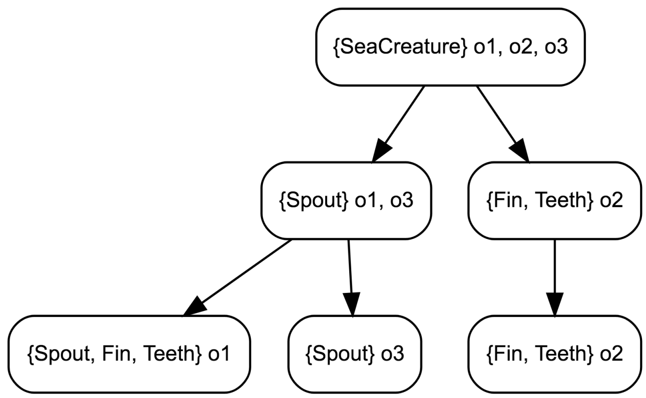
Figure 3: Taxonomy T_1, in tree form and in dataframe form. This taxonomy scores highest of the three, its most inclusive level-2 concept grouping together the two objects (o_1 and o_2) that have the most features in common.



	T2	level	card	n_SPOUT	n_FIN	n_TEETH	Hom	$Incl$	U
1	{SeaCreature} o1, o2, o3	1	3	2.00	2.00	2.00	0.33	1.00	0.33
2	|—{Spout, Fin, Teeth} o1	2	1	1.00	1.00	1.00	1.00	0.33	0.33
3	| °—{Spout, Fin, Teeth} o1	3	1	1.00	1.00	1.00	1.00	0.33	0.33
4	°—{ } o2, o3	2	2	1.00	1.00	1.00	0.00	0.67	0.00
5	|—{Fin, Teeth} o2	3	1	0.00	1.00	1.00	1.00	0.33	0.33
6	°—{Spout} o3	3	1	1.00	0.00	0.00	1.00	0.33	0.33

level	1	2	3	T_2
U	0.333	0.166	0.333	0.277

Figure 4: Taxonomy T_2. This taxonomy scores lowest, the objects in the most inclusive level-2 concept (o_2 and o_3) having no features in common.



	T3	level	card	n_SPOUT	n_FIN	n_TEETH	Hom	$Incl$	U
1	{SeaCreature} o1, o2, o3	1	3	2.00	2.00	2.00	0.33	1.00	0.33
2	|—{Spout} o1, o3	2	2	2.00	1.00	1.00	0.33	0.67	0.22
3	| |—{Spout, Fin, Teeth} o1	3	1	1.00	1.00	1.00	1.00	0.33	0.33
4	| °—{Spout} o3	3	1	1.00	0.00	0.00	1.00	0.33	0.33
5	°—{Fin, Teeth} o2	2	1	0.00	1.00	1.00	1.00	0.33	0.33
6	°—{Fin, Teeth} o2	3	1	0.00	1.00	1.00	1.00	0.33	0.33

level	1	2	3	T_3
U	0.333	0.277	0.333	0.314

Figure 5: Taxonomy T_3. This taxonomy scores between T_1 and T_2, the objects in the most inclusive level-2 concept (o_1 and o_3) having only one property in common.