Concept Utility

Abstract

In 2006, the International Astronomical Union revised their concept PLANET, excluding Pluto where it had before been included. In doing so they insisted that they had improved their concept by revising it. But what could it mean for a concept to be improved? Here we draw on the theory of epistemic utility, which explores how some beliefs are more useful than others, to develop a notion of 'concept utility'. We show how the reliability and informativeness of beliefs, two features that contribute to the utility of a belief, have direct correlates in the concepts that compose our beliefs. 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 these measures can provide us with a single measure of concept utility. The resulting notion of concept utility be used to decide how best to conceptualize an environment, and can rationalize practices of concept revision.

Introduction

One feature of concepts is that we appear to sometimes discover that we have applied them mistakenly – that we have extended them to objects they should not be applied to, or failed to apply them to objects they ought to include. A recent example is a 2006 resolution of the International Astronomical Union. Although they had up to then extended the concept PLANET to Pluto, they decided that given an improved understanding of our solar system, Pluto did not count as a planet after all. It would appear that they found a better way to employ the concept PLANET. But what could make one application of a concept or set of concepts better than another?

Here we tackle this question from the perspective of epistemic utility theory, a branch of epistemology that aims to describe what makes our beliefs useful. In this literature, two elements of beliefs are widely regarded as fundamental to their utility - namely their plausibility, or how likely they are to be true, and their informativeness, or how much they tell us about the world (section 1). We show here that concepts, as the components of beliefs, have two properties that directly correlate with the plausibility and informativeness of beliefs, namely their homogeneity and their inclusiveness. We provide measures for these aspects of concepts, and argue that in combination they allow us to determine a measure of concept utility (section 2). The resulting account allows us to directly compare the utility of competing conceptual schemes, and to rationalize practices of concept revision, as we illustrate by exploring the redefinition of PLANET (section 3).

1 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, it would seem, the measure of utility of a belief or theory is its *plausibility*.

If we adopt the standard Bayesian approach to thinking about the plausibility of a hypothesis, then 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 sole aspect we are concerned with when we decide what to believe. If the only measure of the value of a belief were its plausibility, then we would have little explanation for the kinds of beliefs we are inclined to commit to. Rather than only being interested in acquiring beliefs that are likely to be true, we are also concerned to acquire beliefs that are *informative* – that tell us something substantive about the world by eliminating various possibilities. 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 as some power of the distance between that of the* $2 + 1/50th$ *and that of the* $2 - 1/50th$*, and there is no reason to think that it differs at all from the inverse duplicate ratio.*
>
> 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.
>
> Or he could have made his conclusion conditional, as in $H''c$:
>
> ($H''c$) If the electrostatic force falls off as the nth power of the distance, for some n, then n is between 1.98 and 2.02.
>
> Both $H'c$ and $H''c$ are 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)

If our priority in deciding what to believe were 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 interested in informative propositions, the value of a belief is measured not just by its probability of being true, but by how much it tells us about the world. We value not just plausibility, but also informativeness.

How is informativeness measured? One way is to compare hypotheses in terms of the number of possibilities they exclude. 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 if confirmed would be again more informative. As we can see, the more possibilities a hypothesis rules out, the more informative it is.

As a result, 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 co-vary directly 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:

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

Of course informativeness as such cannot be what we are concerned with, since false beliefs are not useful to us either. 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). Following Huber (2008), we call this the 'informativeness-plausibility' theory of acceptability. What matters about the theory, however, is that these two 'virtues' of belief push in opposite directions. The more possibilities a belief rules out, the more informative it has the potential to be. The fewer possibilities a belief rules out, on the other hand, the more likely it is to be true. And so preferring to adopt hypotheses that are likely to be true pushes us in the opposite direction of preferring hypotheses that are likely to be informative.

How do we decide what to believe, then – the more informative, or the more plausible of our hypotheses? Since we value both, a natural assumption is that we should adopt the most informative hypothesis that meets our tolerance for plausibility in a given context. In a scientific context, our demand on plausibility might be very high; while in another context, it might be much lower. There are doubtless many points of debate and fine-tuning that could be explored further on this question, but that is not our purpose here (for further explorations see Levi 1967, Maher 1993, Huber 2008). Rather, we wish to explore, assuming that something along these lines is right, what follows from these considerations for how we might think of the utility of the components of our beliefs – our concepts.

2 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 probable. For example, given a high probability for H_{al}, we get an even higher probability for H'_{al}:

(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}, we should endorse the former since it is more informative.

But now notice 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 made about. That is, by altering the concept over which we project our inductive generalization. First, we can see that if the range of the concept over which the generalization is projected is narrowed, we increase the plausibility of the hypothesis:

(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 greater range:

(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 can use tools as well as the ‘Alalā , etc. It is clear why Rutz et al. do not embrace H'''_{al}: our evidence about the ‘Alalā studied, coupled with our beliefs that not all crow species are cognitively the same gives H_{al} a high probability, 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 varying the range of 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 target-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, Hooded Crows, New Caledonian Crows, 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*. We define inclusiveness, then, as the proportion of objects in a taxonomy that a concept extends to (see Appendix 1, Definition 2).[1] A generalization extended to a highly inclusive concept will be very informative, and one extended to less inclusive concepts, less informative. This gives us a first principle of concept utility:

[1]The term “inclusiveness” comes from Rosch et al. 1976, where it is used in that sense, see Corter and Gluck 1992.

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 generalizations goes up as the range of things the generalization 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 Ravens, since we know that Ravens are different in many respects from ‘Alalā. And while we decrease the variation within a category, the likelihood of discoveries about one object falling under the concept extending to others increases. Since all ‘Alalā are much more similar than all Corvidae, a discovery about one ‘Alalā is more likely to apply to other ‘Alalā than other Corvidae.

The second feature of a concept that affects the utility of generalizations involving it is therefore what we might call its *homogeneity*. We define homogeneity as the extent to which members of a concept share features (see Appendix 1, Definitions 3 and 4, for a more precise definition). This gives us a second principle of concept utility:

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

Granted that we value both informativeness and plausibility in our beliefs, similarly we will value both inclusiveness and homogeneity in our concepts. And just as informativeness and plausiblity 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 would, see Arnauld & Nicole 1662). Since we value both informativness and reliability in our beliefs, we value both inclusiveness and homogeneity in our concepts. And so 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.

We now propose that this maximizing concept utility as defined here can guide us in both the determination and revision of a conceptual scheme.

Concept determination concerns the “static” problem of dividing up a set of objects into various subsets. Given a domain of objects, and a set of relevant features whose distributions is known relative to those objects, we are interested in partitioning the domain of objects into various categories relative to those features - ‘determining’ how the domain should be ‘conceptualized’.

Concept revision on the other hand concerns the "dynamic" problem of revising a conceptual scheme given new discoveries about an environment. In the next two sections, we proceed to illustrate both of those aspects, first by looking at some toy examples, and then at a natural case.

3 Determining and revising 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. The question we are interested in is how the objects are going to be clustered into distinct categories, relative to that set of properties.

	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 blow-hole (F_1); some have a dorsal fin (F_2), and some have teeth (F_3). Is there an optimal way to partition this group? We could conceptualize them as just one kind of thing, grouping all three objects under one concept. Or, we could think of them as three different kinds of thing – assigning a distinct concept to each object. 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 grouping together o_2 and o_3 (P_{22}), or o_1 and o_3 (P_{23}). In total, we therefore get five conceptualizations or partitions of the domain, representable as follows:

$$\begin{aligned} P_1 &: \quad o_1 o_2 o_3 \\ P_{21} &: \quad o_1 o_2 | o_3 \\ P_{22} &: \quad o_1 | o_2 o_3 \\ P_{23} &: \quad o_1 o_3 | o_2 \\ P_3 &: \quad o_1 | o_2 | o_3 \end{aligned}$$

Which one should be adopt? By calculating the twin values we have identified that feed into the utility of concepts, their inclusiveness and their homogeneity, we can see that one conceptualization emerges as optimal. In effect, it gives us a way to answer the question "how many kinds of thing are there?" 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 the partition $P_{21} = \{C_1 = \{o_1, o_2\}, C_2 = \{o_3\}\}$. What is the utility of this partition? First we calculate its inclusiveness $Incl(P_{21})$. We compute this as the average of the inclusiveness of the concepts in the partition - the inclusiveness, again, simply being the proportion of objects in the domain that the concept extends to. C_1 includes two of the three objects in the domain, so it gets a value of inclusiveness of 2/3. C_2 includes one of the three objects, and so it scores 1/3. The average inclusiveness for the two concepts is 1/2, and so we find that $Incl(P_{21})$=1/2.

Next we calculate the homogeneity $Hom(P_{21})$ of the partition, which we once again treat as average of the homogeneity of the concepts. A measure of the homogeneity of the concepts is simply a measure of the extent to which the objects falling under the concept are similar with respect to the properties that occur in the domain. We measure this as the proportion of objects within a concept that feature a property or lack it, whichever is bigger – assuming that having a property is just as much grounds for regarding two things to be similar as lacking a property. By measuring things in this way, the resulting value will always be a 1/2 or more. However, the minimal value on a scale of should ideally be represented as zero, so we rescale the homogeneity values so that 1/2 is represented as 0, 1 is represented as 1, and other values fall in between (this is done by multiplying the value by (2x-1), see Appendix 1).

Let's see how this works by evaluating the homogeneity of the the C_1 with respect to feature F_1. This feature is possessed by exactly 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 – 1/2, and rescale it. This gives us a value of 0, meaning that $Hom(C_1, F_1) = 0$. For the second property, $Hom(C_1, F_2)$ scores 1, since the property F_2 is shared by all objects in C_1. And that is the same for F_3. The homogeneity score for C_1 is the average of those three, that is $Hom(C_1) = (0 + 1 + 1)/3 = 2/3$. C_2 consists of only one object, so it is maximally homogeneous relative to each property, and scores 1.

We can now combine the scores for inclusiveness and homogeneity to find a utility measure for each concept. For C_1 we find: $U(C_1) = 2/3 \times 2/3 = 4/9$; and for C_2 we find: $U(C_2) = 1 \times 1/3 = 1/3$. The average of the two gives us a utility score for this partition, which is 7/18.

Consider for comparison the partition P_1 consisting of a single concept C encompassing all of the objects. We find that $I(C) = 1$, $Hom(C, F_i) = 1/3$ for each i, and so $Hom(C) = 1/3$. 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 one object that has very 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 the homogeneity of the concepts results in it getting a higher overall score. In fact, P_{21} beats all of the other partitions:

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

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 property 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 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 classifying the objects in a domain, measuring the inclusiveness and homogeneity of different classifications gives us a principled way to choose between them. And hopefully it is now easy to see that maximizing the combination of those values we maximize the utility of the beliefs we use these concepts to form. To see this, 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 properties. Now suppose that you make a new discovery about one member of the group denoted as o_2. You notice that these creatures have a pentadactyl bone structure in their fins. To make a prediction about your environment, you want to project this property onto other members of the group – you form an expectation about other objects in your environment that they might have pentadactyl limbs given that the one you have observed has this feature. Over which individuals do you project this generalization?

This will depend on which conceptualization you have adopted. If you have adopted the second of the two-concept conceptualizations, you might generalize the discovery over all the members of the concept to which you have assigned o_2. In the case of P_{22}, this means you generalize over other individuals denoted under o_2 and also those denoted by o_3. This is a relatively informative inference, telling you about 200 creatures. But because under the second concept in P_{22} you have included creatures that are very dissimilar, failing to share any properties considered so far, we might expect such an inference not to be very reliable (see section 2). If we compare the same generalization made in P_{21}, where the objects denoted by o_2 were grouped together with those of o_1 who are much more similar, your generalization will be far more reliable, but just as informative. And so we can see how optimizing the product of inclusiveness and homogeneity in the concepts we assign to a group in turn optimizes the informativeness and reliability of generalizations we might make in that domain: to optimise the epistemic utility of generalizations we will likely form over a domain, we must first optimize the utility of the concepts we have used to conceptualize that domain.

3.1 Revising a Conceptual Scheme

So much for the determination of an original taxonomy. 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 *properties* in an environment can justify such a revision. This is intuitive: given closer examination of objects in our environment, we can find out that objects that appeared closely related at a glance are actually quite different; or we might find that objects that appeared very different initially turn out to have more in common than we realized. Given such discoveries, we may find ourselves thinking that we need to revise our conceptual scheme. Such considerations echo Waismann's 1945 reflections on what he called the 'open texture' of concepts, which he supposed was needed to accommodate the discovery of new features or dimensions. To illustrate how new features can affect a conceptual scheme, consider what happens if we add two properties, 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 skin color – some are black and white, some are grey; and also feeding habits – some eat krill, others don'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 properties in mind it turns out that o_1 and o_3 have more in common than either has with o_2. 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. The ordering is now: $U(P_{23}) > U(P_{21}) > U(P_{22})$. When new properties are discovered, in other words, the norm of maximizing concept utility can justify revising the scheme.

The optimality of a taxonomy can also be altered by discovering *new objects* in a domain, without discovering any new properties. 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 but is similar with respect to the third. In this case, it might be intuitively unclear whether to think of these four objects as all being the same kind of thing, in which case we would group them together under one concept; or as being two different kinds of thing, grouping the first three together, and separately from the fourth.

First consider the simple partition P_1 into one concept $C_1 = \{o_1, o_2, o_3, o_4\}$. One can check that $U(P_1) = 2/3$. Now consider an alternative partition P_2 that splits the group into two - grouping the first three objects together under one concept $C_{21} = \{o_1, o_2, o_3\}$, and assigning the fourth to its own concept $C_{22} = \{o_4\}$. The second partition scores

$U(P_2) = 1/2$, which is less than $U(P_1) = 2/3$. In this case, then, it is optimal to think of the objects as just one kind of thing: the cost to inclusiveness of splitting the group in two concepts outweighs the gain in homogeneity.

But now consider what happens if we expand the domain by including more objects, without adding any new properties:

	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_4	0	0	1
o_5	0	0	1
o_6	0	0	1

This time, $U(P_2) > U(P_1)$. The utility of the two-concept partition remains at $U(P_2) = 1/2$, but the utility of the single partition has dropped to $U(P_2) = 1/3$. 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 when the range of objects increases, if we retain the same criteria for inclusion in the concept, fully half of the objects are now distinct from the other half with respect to two thirds of the properties. A concept that was originally quite homogeneous can therefore become lose its homogeneity without any new properties appearing among its members, but simply because new objects are discovered that are identical to one of the 'odd man out' objects that had been included in that concept. Changes in the proportion of objects of different kinds within can therefore motivate revising a conceptual scheme. Next, we turn to the recent revision of the concept PLANET, and argue that this is exactly what happened in that case.

4 The Case of 'Planet'

In 2006, the International Astronomical Union formed a committee to resolve a growing dispute over the meaning of 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 PLANET, excluding Pluto and several other newly discovered celestial objects.

In the work of some philosophers, the decision to no longer call Pluto a planet has been characterized as a terminological debate. Chalmers (2011), for example, presents the debate over whether to count Pluto as a planet essentially as a *verbal dispute*, namely as a question of language, rather than as a question of fact. Among astronomers, on the other hand, the case is thought of quite differently: '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). Our view is that Brown was right – that the inclusion of Pluto in the category 'planet' turned out to be factually incorrect, given that our goal in conceptualizing an environment is to maximize the utility of the beliefs we are inclined to form about it. The norms we have set out so far can now be applied to explain why.

4.1 Rationalizing the revision

The explanatory target before us is to show why it is that before the discovery of the new objects in the Kuiper Belt that shared properties with Pluto, splitting the category PLANET into two groups, one including Pluto and one including the other 8 planets, was not justified; but that once the Kuiper Belt objects were discovered, the move becomes justified. As we shall see, the measures we have introduced assign a higher utility to keeping the nine objects within a single category *before* the discovery, but they assign a higher utility to splitting the category into two *after* that discovery. The distinction between planets and non-planets after 2006 was driven largely by whether an object satisfied a criterion defined by Stern and Levison (2000), which they call a *dynamical criterion* – that a planet is 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). Stern and Levison quantified the dynamical criterion in terms of a specific parameter Λ, whose exact definition we don't need to go into here.

To see this, consider Table 1. Here, objects that satisfy Stern and Levison's criterion or not are distinguished, by 1 and 0. The first nine listed bodies are the 9 planets according to the taxonomy received in 2000 since the discovery of Pluto (1930). Underneath are 5 new celestial bodies discovered by Brown and his team between 2000 and 2005, including Eris. Before the discovery of the five lower objects in the table, the line demarcating the category PLANET falls below Pluto. With the discovery of those objects, it falls above Pluto. What justifies restricting the category to above or below that line, before or after the discovery of the new objects?

First, we can show that relative to the new feature identified by Stern and Levison in 2000, the utility of the partition of the 9 planets into two subcategories was lower than the utility of maintaining a single category for the 9 bodies. That feature on its own did not, in other words, justify creating a competing category to PLANET for just the object Pluto. Assume for simplicity the relevant domain in 2000 to consist of only the first nine bodies, namely the traditional nine planets (omitting the Sun, satellites, etc). Relative to that domain, and considering the Stern and Levison criterion $\Lambda > 1$ as the relevant feature to judge homogeneity, the inclusiveness of a category containing only 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). The utility of the category, as the product of these two, is therefore $1/9$. The utility of having a category including the other

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

8 bodies is 8/9 for symmetric reasons. The overall utility of the two categories at this level is therefore $(1/9+8/9)/2 = 1/2$. By contrast, consider keeping a single category PLANET encompassing those nine objects. Its inclusiveness is 1, since it now includes all the objects in the domain, and its homogeneity relative to the Stern and Levison's discriminant is 8/9, since 8 out of the 9 objects share the discriminant property. Scaled according to our algorithm this assigns a homogeneity score of $2 \cdot (8/9) - 1 = 7/9$. This means that even though Pluto is an oddball in terms of the planetary discriminant Λ, the utility of a partition splitting Pluto and the other planets into two categories is lower than the utility of a partition that includes just one category encompassing all nine objects.

Consider now the expanded domain five years later in 2005, when the community's attention is drawn to Eris and to the other four Kuiper Belt objects discovered by Brown and his team. First we can determine the utility of having a partition that includes the old category PLANET (P) encompassing all the objects including Pluto, and a separate category for the new objects, lets call it N. The inclusiveness of P is 9/14, and its homogeneity again is 7/9. The inclusiveness of the complement category N is 5/14, but its homogeneity is 1 (they all lack the discriminant). From our definitions, it follows that the utility of P is 1/2 whereas the utility of N is 5/14. The utility of that partition of the domain – keeping Pluto with the old planets but creating a separate category for the newly discovered objects – is therefore $(1/2 + 5/14)/2 = 6/14$, or 3/7.

On the other hand, contrast that with the partition that groups Pluto with the newly discovered objects under one category (N) and restricts PLANET to the eight other bodies. For this partition, $Incl(P) = 8/14$, and $Hom(P) = 1$; $Incl(N) = 6/14$, and $Hom(N) = 1$.

So $U(P) = 8/14$, and $U(N) = 6/14$, hence the utility of that partition is $1/2$, which is greater than $3/7$. In other words, once the new objects are discovered, our measures predict that there is greater utility to introducing a partition that separates Pluto from the other traditional 8 planets and groups it with the other objects; whereas before the discovery of those objects, a single category including Pluto has a higher utility.

In this concrete case we can see that our measures reflect the pattern of decisions as they in fact have unfolded. First, we can see that the introduction of the Stern Levison feature was not by itself sufficient to force a revision of the received taxonomy before the discovery of the new objects – even though Pluto was distinct from the other planet on this measure. However, once sufficiently many relevant objects had been discovered relative to that feature, the revision is justified. The case of PLANET therefore follows the pattern discussed at the end of the last section, where an anomalous subcategory, when its population grows due to the discovery of new objects rather than properties, can prompt the revision of the conceptual scheme. Once we understand the relevance of both homogeneity and inclusiveness to a taxonomy, due to their role in supporting the utility of the beliefs we are in a position to form using that taxonomy, the decision becomes transparently justifiable.

Admittedly, our analysis simplifies the complexity of the original case, since many more features should be taken into account to calculate the utility of those concepts and of the associated taxonomies (see the discussion in Appendix 2). However, Stern and Levison's dynamical criterion is in a sense the main criterion used by the IAU to delineate between Pluto and the other planets. Similarly, our analysis 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 a contraction of the old category PLANET. 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.

5 Conclusion

Ordinary practices of concept revision suggest that concepts are the kinds of things that we can discover we are employing suboptimally. To make sense of this, it is necessary to have a theory of concept utility - a theory that tells us what an ideal conceptualization of a domain might be. Standard theories of concepts in philosophy offer no such account. The current paper rectifies this by offering an account of concept utility grounded in the rich philosophical notion of epistemic utility, and we believe represents a significant advance in our understanding of both concepts and the reach of epistemic utility theory.

References

Arnauld, A. & Nicole, P. (1662). *Logic or the art of thinking.* Cambridge University Press, 1996.

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

Burge, T. (1979). Individualism and the Mental. In French, Uehling, and Wettstein (eds.) *Midwest Studies in Philosophy IV*, University of Minnesota Press: 73-121.

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

Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press.

Chalmers, D. (2002). The Components of Content. In *Philosophy of Mind: Classical and Contemporary Readings*, D. Chalmers (ed.), Oxford: Oxford University Press, pp. 60833.

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

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

Égré, P. (2013). What's in a Planet? In M. Aloni, M. Franke and F. Roelofsen eds., *The Dynamic, Inquisitive and Visionary Life of* $\phi, ?\phi$ *and* $\Box\phi$*, A Festtschift for J. Groenendijk, M. Stokhof and F. Veltman*, ILLC.

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

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.

Jackson, F. (1998). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*, Oxford: Oxford University 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.

Medin, D. and Schaffer, M. (1978). Context Theory of Classification Learning. *Psychological Review*, 85 (3): 207-238.

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

Murphy, G. and Medin, D. (1985). The Role of Theories in Conceptual Coherence, *Psychological Review* 92 (3): 289-316.

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., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, 8(3), 382-439.

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.

Searle, J. (1983). *Intentionality.* Oxford University Press.

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).

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

Appendix 1

This appendix lays out some definitions intended to make precise the notion of concept utility. To keep things simple we assume a finite domain D of objects. We regard concepts as intensional entities, but in what follows, without loss of generality, we handle concepts extensionally and identify them with subsets of the domain (see Definition 1). We take for granted the notion of partition (a set of concepts that are mutually exclusive and exhaustive of the domain). We say that a partition refines another partition if each concept of the former is a subset of a concept of the latter. Similarly, we take for granted the notion of a feature, namely a binary property that an object can have or fail to have (see the Illustration below).

Definitions

Definition 1. *A concept C is a subset of the domain D (i.e. the set of C-objects, or objects satisfying C).*

Definition 2. *A conceptual scheme or taxonomy is a finite family $(P_i)_{i \leq m}$ of partitions of the domain into distinct concepts, such that for each i, the partition P_{i+1} is a refinement of P_i. The level k in a taxonomy is the corresponding partition P_k.*

Definition 3. *The inclusiveness of a concept C, noted $Incl(C)$, is the proportion of objects of D satisfying C.*

Definition 4. *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 .5 and a maximum value of 1 (when the modal proportion is x, the homogeneity is $2x - 1$).*

Definition 5. *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 6. *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:*

$$U(C, (F_i)_{i \leq n}) = Incl(C) \times Hom(C)$$

Definition 7. *The epistemic utility E of a level within a taxonomy is the average of the epistemic utilities of concepts at that level.*

Definition 8. *The epistemic utility of a taxonomy is the sum of the epistemic utilities of the levels of the partition, divided by the number of levels.*

Remark 1. *We could define the epistemic utility of a level to be the product of the average inclusiveness of concepts at that level by the product of their average homogeneities. However, that definition appears slightly less natural to us, if indeed epistemic utility is first attached to concepts.*

Remark 2. *We could assign different weights to different features. We assume equal weights in what follows, but the generalization would pose no difficulty.*