# How do people use and appraise concepts?

James A. Hampton

Professor Emeritus, Department of Psychology,

City, University of London.

*This chapter is based on a presentation to the Conceptual Engineering Online Seminar, CEOS21 St Andrews University. The research described was supported by the European Union's Horizon 2020 research and innovation programme, grant agreement nos. 681422 and 742204.*

To approach the many challenges involved in the notion of engineering concepts, it is important to have a clear idea of the starting point – the concepts that people use in their everyday lives, in conversations and in expressing beliefs, desires, intentions and so forth. The first Section of this chapter introduces evidence that I have accumulated over the last many years concerning the flexibility, context-dependence, and vagueness of such common concepts. The concept engineer needs to understand the structure of the "raw material" with which she is working.

In Section 2, I summarise some results concerning the amount of individual variation that exists within a single relatively homogenous population of individuals, both in what examples are considered most typical of a category, and in what properties are considered most central to a category (Hampton & Passanisi, 2016). The demonstration of a significant degree of reliable variation in responses to these questions, that is stable over time, illustrates the potential for conceptual engineering interventions to affect individuals' concepts. Concepts are rarely fixed or uniformly understood in the same way. I then discuss the different ways in which concepts track the world, and I speculate about how different conceptual domains are governed by different criteria of validity or acceptability. This reflection will lead to Section 3 in which I summarise some results from recent research (Thorne, Quilty-Dunn, Smortchova, Shea, & Hampton, 2021), investigating for the first time how people view the value of different familiar concepts. Knowing how people appraise concepts will be a key to developing ideas for improving the reliability and validity of the concepts that they commonly use. A final section contains some (largely speculative) discussion about the implications of work in psychology of concepts for the project of conceptual engineering.

## 1. Concepts for Every Day

Each of the substantive words in a language – that is principally the nouns, adjectives, verbs, and adverbs – has an associated meaning and labels an associated concept. Without going into the complications of how to differentiate meanings and concepts, in what follows I will be describing empirical research into the mental

representation of the concept to which a word refers. The word "sport" thus refers to the concept of SPORT, and its mental representation will comprise knowledge of the different activities (e.g. tennis, football) that people consider to be sports, and knowledge of the common properties of such activities (e.g. competitive, leisure-based) which people consider render them suitable for membership of that category. I will refer to these as mental representations as "concepts" in keeping with practice in psychology.

A key finding from research in psychology over the last 50 years, is that concepts are adaptive and flexible (Hampton, 1995; Murphy, 2002; Rosch, 1975). They evolve and change as they serve different cognitive and social functions. They are also often characterized by a degree of vagueness, which can sometimes seem worrying, but in practice can aid social cohesion and understanding. For example, in a set of negotiations between parties with different goals, mediators will look for a form of words which is sufficiently vague that each party believes that it adequately meets their goals (a "fudge"). Being vague also means that concepts can naturally evolve over time, as has frequently happened with the meanings of words. It also allows members of a community to "share" a concept, without individual differences necessarily becoming apparent and creating disagreements. Vague concepts are also key features of political rhetoric – allowing the hearer to create their own idea of what they are being promised. In the infamous referendum campaign for the UK to leave the European Union, key concepts such as "sovereignty" and "control" were used to great effect without any need to explain what achievement of such goals would mean in practical terms.

The simplest way to represent concepts such as SPORT is through the idea of a prototype (Rosch, 1975), or polymorphous concept (Dennis, Hampton, & Lea, 1973; Hampton, 1979; Ryle, 1951 cited in Urmson, 1970). As described above, there are two aspects to our knowledge of any concept, which for convenience we will call the "extension" and the "intension" (different research disciplines have other ways to define these terms). The extension is the class of things in the world that "fall under the concept" and so are appropriately labelled with the category name. For SPORT this would include football, tennis, and athletics.The intension is the set of properties (also called attributes or features) which are associated with the concept. Again for SPORT, this set would include "is competitive", "requires training" and "involves physical exertion". Prototype theory of concepts simply proposes that the concept is represented as a set of properties, each weighted according to its diagnosticity for categorization, together with a threshold criterion (Hampton, 1995). To belong in the extension, an item must possess a sufficient weight of the right properties to pass the threshold. In effect the prototype model proposes that concepts are linear classifiers.[1]

Notice how this simple account possesses many of the right characteristics to explain the evidence of concept vagueness. Where an item has some properties but not all, it will fall close to the category boundary, meaning that people will be uncertain whether it belongs or not (take for example, chess or frisbee as sports). There is plenty of scope for individual variation in how weights are given to different properties, without major disagreements arising too often. There is also no difficulty in the evolution and change of

[1] It is important to note that while concepts have prototypical instances, those instances are not themselves prototypes according to current prototype theory, which treats prototypes as sets of correlated features.

the concepts, as new experiences can lead to new properties becoming associated with the concept (as in the advent of computer games as potential sports), and others being dropped (as in the Victorian idealised notion of the amateur sportsman, Cronin, 2014)

As should be immediately obvious, there are also concepts for which this simplistic scheme does not provide a good account. In particular, for natural kinds such as biological or chemical terms (TIGER or GOLD), appearance is often secondary to a causal "essence" in determining category membership (Gelman, 2004; Medin & Ortony, 1989 ). In a fantasy context it may even be considered that appearance is irrelevant (Rips, 1989). Other cases may include a causal schema in which possessing correct relations among the properties are also key parts of the concept (Ahn et al., 2001; Murphy, 2002; Rehder & Kim, 2010). A car doesn't just have a motor and wheels and run along a road, it is also important that the motor causes the wheels to turn so that the car moves along and that the wheels are of the right kind for running on a road. Each part of a complex artifact like a vehicle has evolved through a design process to provide a close fit to the function of other parts. In learning studies with artificial categories, having the correct causal connection can have a marked effect on whether an example is considered a category member (Rehder & Kim, 2010).

While it is clearly true that our conceptual knowledge is not a simple set of linear classifiers using sets of features to divide the world into categories and label them, it is also important to note that deeper conceptual knowledge is often patchy and poorly understood. The classic study by Rozenblit and Keil (2002) on the Illusion of Explanatory Depth nicely showed how people believe that when they know how to use something, they think that they also know how it works. In fact, this confidence is often completely unjustified. In this regard, a very nice study by Lawson (2006) provides a good demonstration. Her participants were shown the outline figure of a bicycle seen in Figure 1, and asked to add in lines for the frame, the pedals and the chain. Ability to draw the correct arrangement was markedly poor – even in the case of participants who used a bicycle daily. The reader is invited to take the challenge!

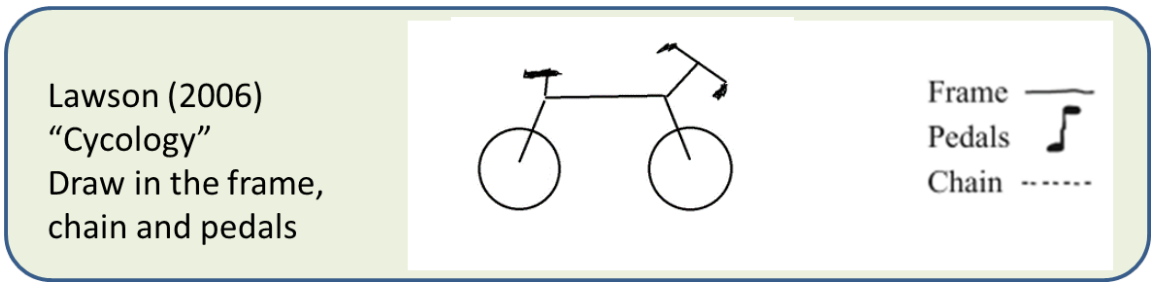


Figure 1: the outline of a bicycle to which the frame, pedals and chain need to be added (Lawson, 2006) (the outline has been redrawn from the original).

One important line of evidence concerning the prototype nature of many concepts comes from studies I have conducted over the years on logical propositions involving such concepts. Starting with Hampton (1982), I demonstrated that the semantic relation of superordination "A chair is a kind of furniture" is (surprisingly) not one of class inclusion. While everyone agrees that the statement is true, (chairs are one of the most typical kinds of furniture), many will also acknowledge that there are exceptions – for example a child's

car-seat is a chair, but not furniture. It appears that "A is a kind of B" does not necessarily imply that all As are Bs. Rather, the sentence can be considered as a generic construction, like "birds can fly" or "cars have four wheels" – a statement of what is normally expected to be the case. In fact, Wierzbicka (1984) argues for a multiplicity of different semantic relations that can be expressed by the same "is a kind of" sentence form, only some of which are transitive.

Other evidence along similar lines comes from studies I conducted of conjunction, negation, and disjunction of common category concepts (Hampton, 1987; 1988a; 1988b; 1997). Again, the expected logical form of the phrases does not fully correspond to how they are used. For example, in my studies 54% of a sample agreed that an elevator (a lift in the UK) is a vehicle. Yet 80% of a sample from the same population agreed that it was a machine that is a vehicle. The implication is that a sizeable number would agree that it is both a machine and a vehicle, but not allow that it was a vehicle as such. Placing the concept of vehicle in the context of machines changes the concept in a way that makes an elevator a better fit.

For negation, an example from my study was as follows. Asked if a desk lamp is a type of furniture, 80% agreed, but in a different condition 74% agreed that a desk lamp is a household appliance that is NOT furniture. Once again, the context changes the basis for categorisation. Finally, similar results can be shown for disjunction. In my study, 50% of the sample agreed that a mushroom is either a fruit or a vegetable, even though only 10% of another sample agreed it was a vegetable, and no-one would allow that it was a fruit[2].

The account that I offer for these rather disconcerting results is that when people are asked to judge superordination, conjunction, negation or disjunction of categories, rather than thinking extensionally about set overlaps or set unions, they instead retrieve the two sets of properties associated with each concept and proceed to combine them into a composite concept[3] (Cohen & Murphy, 1984; Hampton, 1991). As Hampton (1982) concludes in discussing the car-seat/chair case:

> "… subjects do not apparently consider counter-examples as disconfirming the truth of unquantified category statements. This result implies that such statements rely for their verification on the similarity of descriptive meaning of the two concepts (their intensional meaning) rather than on the inclusion of one class of things within another class (the extensional aspect of concepts)."

In the case of a conjunction, to belong in <u>both</u> sets, an example needs to meet <u>both</u> criteria for categorization – that is, in a standard classical model, it needs to possess all the features in the *union* of the two sets of properties. But prototype properties are not all necessary for membership, and furthermore there may be conflict between the properties of one concept and those of another. Take the example of the elevator. Vehicles normally travel on roads or rails and typically take you from one place on a map to another. Elevators lack either of these features, though they do have the important feature of using energy and

[2] Ironically enough, a mushroom is in fact the fruiting body of a fungus.

[3] This failure to think extensionally may explain the difficulty many students experience with Introductory Logic courses.

mechanical force to transport people and goods from one location to another in a way that can be reversed and reused at will. Elevators are thus borderline vehicles, as seen in the 54% endorsement of their category membership. When composing the conjunctive concept of machines that are vehicles, people will place more weight on the features that are true of both concepts – such as requiring an energy source which is turned into mechanical force and movement. Elevators are a better fit to this set of features, and so their degree of acceptance into the conjunctive category is greater.

A similar account can be given for the other examples. Household appliances are often electrical, and possessing this feature boosts the membership of the desk lamp in the conjunction, to the extent that its possession of other furniture properties (convenience, mobility) can be ignored when evaluating it as a Household Appliance that is NOT furniture. The case of the mushroom can be explained by suggesting that when two conceptual categories are close coordinates such as is the case for fruits and vegetables, and even show some overlap (tomatoes are judged to be in both categories to a degree), then a composite concept can be formed of the disjunctive set from the properties that the two concepts have in common (i.e. the *intersection* of the two property sets). In this case that would involve properties such as edible, vegan, cultivated, found on the fresh produce aisle in the supermarket, all of which provide a good match with mushrooms.

Figure 2 illustrates how prototype composition handles a classic case in the philosophical debates about concepts and their compositionality (Fodor & Lepore, 2002; Hampton & Jönsson, 2012). According to a classic paper by Osherson and Smith (1981), typical pet fish are neither typical pets nor typical fish, and it is therefore not possible to explain how typicality can be mapped from individual conjuncts (pet and fish) onto the conjunction (pet fish). By examining the properties of the concepts however (Hampton, 1987) one can see how a concept conjunction can be created with the right properties. In the figure the features of fish are divided into those that can also be true of pets, and those that are not true of pets. Similarly, the features of pets are divided into those that can also be true of fish, and those than cannot. Consequently, examples like GUPPIES can match all the features of the conjunction, while mismatching individual features of either conjunct.

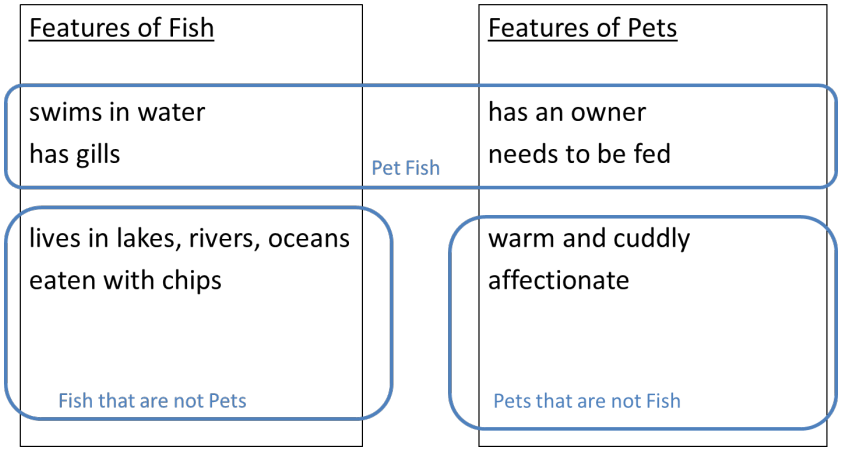

Figure 2: How prototype compositionality explains the "pet fish" problem

## 2. Individual variations in concept representation

Thus far, I have been concerned to illustrate how empirical data on people's understanding and use of concepts can be well explained by assuming that at least some concepts are primarily represented as intensions rather than the "externalist" view that sees them as atomic labels that simply point to extensions (Fodor, 1998; Fodor & Lepore, 2002). A common argument for the latter view is the question of how it can be said that two people are talking about the same thing. If they have individually different concepts underlying a word's meaning then surely they must be talking past each other. Aligned with that problem is the question of how it would be possible for someone to have the "wrong" concept, if the meaning of the word for that person simply is the set of properties that they personally associate with that word.

I will return to these two issues later in this section, but first I want to present some empirical evidence on the question of whether individuals actually *do* have different conceptual representations of any given concept. This research was conducted with Alessia Passanisi at Kore University in Enna, Sicily (Hampton & Passanisi, 2016). Early work by Barsalou (1989) investigated just how stable people's perceptions of prototypes are. He studied a common task used to investigate prototypes – a rating of the typicality of different category members. Because members of a prototype category may possess different degrees of match with the properties, those that pass the threshold for membership will still differ in how well they fit the concept. This gives rise to a sense that some members are more typical or representative of the class than others (Rosch, 1975). Chairs and tables are more typical of furniture than televisions or paintings, as they possess more of the typical functions and appearance of furniture. When they examined individual responses to the task, Barsalou and colleagues found that one student's typicality judgments correlated at only .50 with another's from the same group. Of course, this low correlation may simply reflect noise in the rating process, but when students did the same task on two different occasions, the correlation between the two sets of ratings fell from .92 after one hour to .76 after a month, strikingly above the average between-individual correlation.

The fact that within-individual correlations are higher than between-individual correlations shows that there is a degree of stable individual variation in considering what items are typical in a category. In the studies with Passanisi, we confirmed this finding, but at the same time we looked at category *intensions.* An equivalent task for category intensions is to ask participants to rate how *important* each property is, as a part of the concept's definition. In the task, they were told to "tell us how important you think each feature is for deciding whether something is in the category". A rating scale of 1 to 7 was used. In each of several studies, participants performed the rating task twice with an intervening period of a week. The same participants also rated the typicality of category members at the same time. Thus, for any individual we had their typicality and importance ratings on week 1, and the same ratings given again on week 2, to the same set of categories. All the participants were drawn from the same relatively homogenous population of Italian-speaking undergraduate students studying at the Sicilian university.

Our first interest was to show that, like typicality ratings, ratings of feature importance showed agreement between individuals, and consistency across time. In this analysis we found for both tasks that there was a fair degree of consensus (average between-individual correlations of around .4 to .5), and a higher degree of within-individual consistency across weeks (correlations of around .6 to .8). Importance ratings had rather lower correlations than typicality. It therefore appears that stable individual differences can be shown not only in extensional typicality ratings, but in intensional property importance ratings.

The most important finding of our research however emerged when we looked at the relation between the two types of rating task. A person's set of judgments on any task could be used to calculate their similarity to other people in the group using a simple correlation. These correlations were then put into a second level of analysis to look at how similar all pairs of individuals in a group were, (a) in terms of typicality judgments and (b) in terms of importance judgments. We expected to show that if two individuals had a similar idea of what were the most typical sports (e.g. triathlon and marathon) they would also share an idea of which features were most important in making something a sport (e.g. training and fitness). But this was not the case. Across a set of studies, we repeatedly failed to show any correspondence between those people who showed high similarity for typicality and those who showed high similarity for importance. At this point it is worth pointing to a subsequent study (Djajal, Hampton, Storms, & Heyman, 2018) which came to a different conclusion, albeit with a different methodology – category membership judgments were used rather than typicality for extensions, and both property generation and importance ratings were used for intensions. Djalal et al. used an object-by-feature applicability matrix (De Deyne et al., 2008) applied to a person's generated or rated properties to predict either their own categorization behaviour or that of another randomly selected participant. The results showed clearly that the properties that an individual generates or endorses were better at predicting their own category judgments than those of others.

What is the relevance of these results to the current chapter? The results serve two functions. The first is to emphasize the reality of individual variation in concepts, or at least in how people report on the content of their concept representations. Results for both extensional and intensional tasks show that people have their own consistent views which may differ systematically from others in their cohort or language group. It is not possible therefore to dismiss weak correlations between people's judgments as just reflecting a lack of meaning or interest in the tasks. The second is the possibility that intensions and extensions are not as tightly connected in semantic memory as has usually been assumed. How could it be that people who have similar views on typicality do not have similar views on feature importance? One possibility concerns the sources of information from which we build our concepts. In Hampton & Passanisi (2016) we suggest that, when judging typicality, people may reflect on their experience with the world around them – how frequently do they see different objects or activities in the context of a given category? How closely do they associate the category members with each other? Extensional information represented for a concept may be directly influenced by types of experience that people have with the world, so that similar typicality ratings would reflect similar experiences and memories. On

the other hand, when considering which features are more strongly associated with a concept, people may reflect on the concept in a very different way. They may, for example consider the consequences of a feature being missing on the likely presence of other features. Birds that do not fly will have small wings, will not perch in trees, and will not migrate for the winter. Considering this web of belief, whereby features are not independent, but each feature can cause or enable others (Quine, 1970), provides a possible source of information for judging property importance that is independent of individual lived experience, and more dependent on reasoning processes applied to encoded knowledge. To summarise, typicality may reflect experience with the category in the world, whereas feature importance may depend on "mini-theories" about how the features fit together into some schematic representation (Barsalou & Hale, 1993). This account is highly speculative, and a satisfactory explanation of the findings will require further exploration of people's intensional and extensional knowledge.

Individuals may have systematically different prototypes, but it should not be overlooked that, of course, they also agree to a large extent. Our results do not challenge Frege's notion of intensions determining extensions (Cappelen 2018). There is still overwhelming evidence that people categorize examples by examining the features that they possess and their relevance to the category. But the results do underline the flexibility in the process, arising from other factors that influence typicality in extensions and that affect judged importance of properties.

## 3. Three sources of concept validation

Given our results, how is it that we can be said to be talking about the same thing, despite individual variations? What would make one say that a person has the "wrong" concept rather than just a "different" concept of some category? I want to propose that the answer lies in considering the different sources of validation that apply to different sets of concepts. By validation I mean providing some normative criteria for determining whether a concept is "correct" or not. These sources of validation relate to (1) the physical world, (2) the social world, and (3) the internal psychological world of experience. Any program designed to shift concepts in a particular direction will need to identify in which of these worlds it is necessary to operate.

The physical world provides us with the natural kind and reference-based concepts that are familiar in science and academic disciplines. To understand this world, we have developed a range of concepts that depend on reality for their validation. The semantics of these concept terms is strictly externalist. We develop concepts such as weight and mass, or heat and temperature because they provide a framework of explanation which is both internally coherent and enables us to predict the behaviour of the external world (Carnap, 1947). Developing these concepts has taken centuries of scholarly progress, and consequently they mostly need to be taught through formal education. When uncertain about a belief or a categorization that uses these terms, people should be willing to defer to the accepted definitions, which have been continually refined ever since the Enlightenment to provide the basis for the simplest, most coherent, and most empirically sound explanations of the physical world.

Alternative beliefs about the nature of the world arise of course with regularity, sustained by groups of people who find them plausible, but in the case of natural kind concepts, it is (in principle at least) possible to appeal to experiment or predictive power to determine which beliefs are right and which are wrong. For example, classifying a whale as a mammal provides a single satisfactory explanation for its need to breathe air, its being viviparous and having mammary glands. As a fish, all of these features would have to be considered as exceptional.

Turning to the social world, concepts are constructed and validated in a different way. Some will depend on cultural history – as for example MARRIAGE or MONEY – and will be open to challenge and reformulation as society evolves. Concepts such as MARRIAGE and MONEY, or CRIME and OWNERSHIP have been codified by the legislation of each nation and can only be changed by the governing authorities. For less consequential concepts, an important source of consensus comes from the need to cooperate with one's language community. If a new term for popular music emerges – garage, grunge, house, hip-hop and so forth – then a speaker who wishes to discuss music is obliged to learn the relevant concepts underlying these terms, which often may be vaguely defined[4]. Having the wrong concepts will lead to a failure of communication and rejection from the conversation. This type of category is widespread throughout daily life – be it genres of book, film, computer game, or types of restaurants, clothing styles or vehicles. Consensus is achieved by the social pressure of needing to be understood and to contribute to conversations. There is no need for any of the categories to have a "deep" meaning beyond some superficial similarity within the class.

Given most people's lack of knowledge of the definition of actual natural kind terms (Putnam, 1975), social processes will also be invoked to validate natural kinds within a language community. In this way, natural kind concepts like FRUIT or WATER can develop alternative non-technical senses, as illustrated by Malt (1994). In her study, Malt showed that people's willingness to categorize a liquid as water was only indirectly linked to their belief in how much $H_2O$ it contained. The social function of the substance in question was equally influential in how the word was used.

Borderline cases of socially validated categories will be common, as will degrees of typicality. It is also in the realm of social concepts that efforts to engineer concepts are likely to be focussed, for example by introducing new terminology to replace stereotypical terms that have acquired pejorative or negative associations. A new term can provide a basis for driving a deeper change in the concept underlying the category concerned. "Service user" or "Someone with mental health issues" can replace terms like "insane" or "mad", in a way that both present someone as an active seeker of services, and as someone who is not defined by their mental health status.

The third domain of concepts, after those tied to physical and social realities, is concepts that are strictly personal. The belief that the world is flat can be demonstrated to be "wrong" or false. Someone who chooses to call blue objects "red" will find communication difficult. But the belief that the taste of chocolate is unpleasant is validated personally. . While one can argue about the taste of chocolate, one cannot deny the other's

[4] Categories of music most probably rely on similarity to known examples, at least initially.

view that they do not like it. Internally based concepts are especially important in how people construe their own feelings, emotions and motivations, but being private to the individual, they may be harder to access[5].

I have presented the three domains as if they are fixed, but in fact concepts can move up and down the hierarchy (if one places natural kinds at the top and personal concepts at the bottom of a scale), and the situation is necessarily more complicated that this simple scheme would suggest. Those at the top level will mostly have begun as socially-validated concepts – as in the case of water before the development of chemistry. Other concepts which were considered as being top-level were subsequently deemed to be invalid (examples would be phlogiston as an explanation of heat, or the humours as accounts of personality.)

The promotion of concepts to natural-kind status is particularly evident in medicine. Advances in understanding infectious diseases or cancerous growths have led to clearly defined, empirically based definitions of medical conditions as natural kinds. Once an aetiology is established for a condition, a category of patients can be clearly delineated, even if symptomatically diverse. At the same time, the development of conventions on medical terminology greatly reduce ambiguity. It is this that leads Burge (1979) to question the status of the belief of someone claiming to have arthritis in their thigh. Many medical concepts have been engineered (like other concepts in science) by seeking natural-kind status for them (that is, providing a causal essence that determines the extension). By creating an international ontology (Welty & Smith, 2001), concept terms become firmly fixed, allowing for highly focussed literature searching and meta-analysis of research studies. Interestingly, in medicine there are also terms such as *pyrexia* which simply replace the more commonly understood word (*fever)* without any change in the concept, but with a greater chance of impressing the patient!

In psychiatry the goal of engineering a set of natural kind concepts remains mostly beyond reach. The issue of psychiatric diagnostic categories is very much a live one at the time of writing (Browne, 2022). Different competing taxonomies exist to challenge the International Classification of Diseases (ICD) issued by the World Health Organization, and the Diagnostic and Statistical Manual of Mental Disorders (DSM V) of the American Psychiatric Association. The distinction between "disease" and "disorder" is telling here, as indicating the dualist confusion of whether poor mental health has a physical or psychological basis (to quote a well-known psychiatrist "we tell the patient there is nothing physically wrong with you, and then suggest they take one of these tablets each day"). Likewise, there is a debate on whether classifications should focus on aetiology of symptoms or take an atheoretical statistical clustering approach based on symptom distributions (see for example Xiao et al., 2023, for a history of how psychiatric terms have evolved over time.)

Another illustrative case of a concept moving between domains comes from astronomy. Discoveries of the variety and amount of orbiting material in the outer parts of the Sun's gravitational field led astronomers to re-engineer the concept of planet and to

[5] According to some readings of Wittgenstein (1956), concepts solely restricted to a private language may be impossible, but the issue is highly controversial (Candlish & Wrisley, 2019).

reclassify Pluto as just one of many hundreds of identifiable "dwarf planets" rather than one of the 9 planets whose names we learned in school. The useful mnemonic *"My Very Educated Mother Just Served Us Nine Pizzas*" has had to be rewritten as "*My Very Educated Mother Just Served Us Nachos".*

**4. Concept appraisal**

In this final section, I present and discuss evidence recently published (Thorne et al., 2021) on how people make judgments about their own conceptual repertoire. When you hear someone use a concept term, what kind of appraisal might you make about the term? Is it a term that you know well and feel relatively well-informed about, or is it a term that you would need to look up or about which you would defer to an expert? Is it a term that you believe can be used for serious debate, or is it just a word that gets used in a wide range of contexts in a mostly vague manner? Is the named category something enduring and with a deep basis, or is it just a vaguely defined notion that does not support further inferences? Questions such as these were addressed in the studies (Thorne et al.). Initial pre-theoretical discussions led us to speculate that there could be two major axes on which concepts would fall. One would reflect the degree to which an individual feels they understand a concept term. I may feel I understand the occupation of police officer very well but have little idea of that of market trader. The second axis was a feeling of how dependable a term was. How informative is it, does it support inductive inferences, is there a field of expertise that you could appeal to for further information? For example, I may feel that leukaemia as a concept has all these properties, whereas having a verruca on one's toe is very much less consequential or informative.

We began by brainstorming a set of eight dimensions which we believed might tap into these different forms of concept appraisal. The method we used was taken from Haslam, Rothschild, and Ernst (2000), who investigated essentialist beliefs about social categories. Haslam et al. identified nine different dimensions expressing aspects of essentialism, and had participants rate a set of 20 pairs of social categories along each dimension. Principal Components Analysis (PCA), a way of reducing the complexity of the full set of correlations between the nine dimensions, showed that the nine dimensions could be reduced to two axes which captured people's essentialist beliefs about the categories. One was Naturalness, incorporating immutability, stability, necessity, and discreteness, and the other was Entitativity – the "thingness" of a category, incorporating its uniformity, informativeness, inherence, and exclusivity.

Using this method, we aimed to condense our eight dimensions down to their underlying axes in a similar way. This was primarily exploratory research. We did not know at the outset whether people would agree on their ratings of categories on the dimensions, nor whether the correlations between dimensions could be captured on a small number of axes as in Haslam et al.'s study. Figure 3 shows the eight dimensions that we used. Each dimension was rated by a different group of around 50 participants, who each rated the same four domains of concepts on their given dimension. The four domains were Occupations, Recreations, Health conditions, and Biological kinds (plants and animals). For each domain we chose 40 subcategories to be rated, selected to vary in terms of how well-known they were and how reliable we thought they may be.

## Eight Dimensions of concept metacognitive feelings

**Explain**. *How confident do you feel that* you could explain the category to someone who knew nothing about the category?

**Accuracy**. *How sure are you that everything you know about the category is true?*

**How Much Do You Know.** *How much do you think you know about the category?*

**Communicate**. *How likely are people to understand each other using these category names?*

**Informativeness**. *How much does knowing some thing belongs to a category tell us about this thing?*

**Induction**. *If you found out that three members of a category have a characteristic how likely is it that other members of the category will also have this characteristic?*

**How Much to Learn.** *How long will it take someone to become an expert in this category?*

**Deference**. *How likely would you be to defer to an expert about this category?*

Figure 3 Eight dimensions of concept appraisal (Thorne et al. , 2021)

On analysis of the data, we found that a PCA solution could reduce the 8 dimensions to two axes of 4 dimensions each. A factor that we labelled Understanding captured the strong intercorrelation of the dimensions *Explain, Accuracy*, *How Much Do You Know*, and *Communicate* (see Figure 3). Well understood categories could be explained, supported accurate beliefs, were well known, and were useful for communication.

The other four dimensions were also strongly intercorrelated and formed what we termed a Dependability factor (*Induction, Informativeness, How Much to Learn*, and *Deference*). Our pre-theoretical intuitions were thus largely supported by the data. On the other hand, when we examined each of the four category domains individually, we found major differences. For all four domains, the four dimensions of Understanding held together as a single factor, but the other dimensions, which we had considered as relating to Dependability, showed quite different patterns across the four domains. For example, for Occupations, *How Much You Know* turned out to load less well on Understanding, and to load negatively on Dependability. This pattern captured the variation between complex professional roles (not well understood but highly dependable) and simple descriptive categories (better understood but less dependable). Health conditions also broke away from the overall pattern. *Induction* and *Deference* were negatively related to Understanding, while the second factor was only represented by *Informativeness* and *How Much to Learn*. In health matters, the less well you understand something the more willing you are to defer to the experts. Meanwhile, plants and animals generated yet another solution with 3 different axes.

A second study using the same method but with four new domains (sports, fruits and vegetables, clothing, and furniture) confirmed the result that while Understanding was consistently captured by the first four dimensions, the other four would create very different patterns of intercorrelation, depending on which domain was being explored. Further studies showed how basic level categories (Rosch et al., 1975) are better understood, but not more dependable, and how variation in Dependability can be used to predict choices in an inductive reasoning task (Patalano & Ross, 2007). For example, participants were told that for two personality types, most police officers were type X while most library users were type Y and were then asked to guess the personality type of an individual who was both a police officer AND a library user. In this task, differences in Dependability were significantly able to predict people's choices, whereas differences in Understanding did not.

## 5. Implications for Conceptual Engineering

In this chapter I have outlined a view of the mental representation of concepts as a highly adaptable and flexible system for everyday reasoning and communication. People have some metacognitive awareness of how well they understand their everyday concepts and can make complex and domain-specific judgments about the inductive potential and explanatory power of different conceptual categories.

Most concepts that we use in daily life are basic-level terms (Rosch et al. 1976) which are broadly learned and kept current through people verbally interacting with others, and through numerous cultural media. Because they are vaguely constituted, they may be readily adapted by the engineer, often for political or social ends. More often than not, however, they most likely evolve along with changes in society. As telephones no longer have dials and are no longer attached by wires or hanging on a wall, so the concept has evolved as the category and the prototype have shifted. The evolution of other concepts can be facilitated by progressive politicians and other influential people. After homosexual acts ceased to be illegal in England and Wales with the Sexual Offences Act (1967), some 46 years later the concept of marriage was changed in law to include same sex partners. At the time of writing the concepts of MONEY and MONETARY VALUE are undergoing a new revolution with the advent of cybercurrencies and Non-Fungible Tokens.

Changing "top level" concepts may be a matter for politics, persuasion, and public information campaigns, together with the development of educational curricula in schools. Clearly, if not based on hard science, these attempts to revise concepts can be used for malign ends as easily as for beneficial ones. The psychological concepts of Intelligence and Personality have divided the field since attempts to use psychometric methods to explicate them in the early twentieth century, and the issue is still far from resolved (Serpico, 2021).

For the "middle level" of everyday concepts like sports and fruits, there is less urgent need for a fixed agreed definition. Only rarely do issues of category membership arise in legal contexts. An infamous case concerned the Jaffa Cake, a confectionary item common in England. Sales tax (VAT) was payable on cakes but not on biscuits (cookies in the US), so a legal ruling had to be made as to whether the Jaffa Cake was a cake or a biscuit (HMRC, 2023). (The tribunal decided that they were cakes, as cakes become hard when stale while biscuits become soft). The concept of sports also came before a panel of judges:

> *At a High Court hearing in 2015, the Aylesbury-based EBU [English Bridge Union] unsuccessfully challenged Sport England over its refusal to recognise bridge. EBU lawyers told the High Court the game was based on rules, fairness and competition like other sports and was available to older people, to whom it brought a sense of inclusion and community.*
> *But the judge said Sport England's current policy defined sport as involving "physical activity" and its move to reject reclassification of bridge was correct. So-called duplicate bridge, a form of contract bridge, is already considered a sport by the tax authorities of Austria, Denmark, France, the Netherlands, and Belgium, but not by Ireland and Sweden. It was classified as a sport by the International Olympic Committee in 1998.* (BBC News, 2017)

The examples show how concepts in the public realm are negotiated for the interests of different parties. Particularly noteworthy is how each side can appeal to the presence or absence of generic prototypical features of the contested example to make their case. Similarly, experts in different jurisdictions can arrive at different answers.

There is an interesting boundary region between concepts that have conventionally agreed definitions, and those which are open to argument and debate. In this respect note that when concepts are enshrined in law, there is then (at last in UK common law) a transitional period in which case law can be established. Just as the concepts of CHAIR or PHONE change with the appearance of new designs, so the interpretation of concepts in law has to be "bedded in" with the experience of applying the law to different cases. Making a legal case requires a search for relevant case-law to support your side, just as when classifying some novel food as a fruit, people look for its similarity to known fruits (Smits et al., 2002)

I have said little about the third level of my concept hierarchy, but I suspect that to the extent that a person can have a personal idiosyncratic concept which is not subject to constraints from the language community or from the facts of the world, then it is unlikely to be susceptible to engineering in the same way as other concepts. In the course of psychotherapy, it may be that the personal conceptual world of the client may be open to explication by the therapist in a way that leads the client to a better understanding of their behaviour and emotions through the development of better concepts (McLeod & Sundet, 2022). Seeing the process of therapeutic intervention in the context of conceptual engineering may be a way to describe and analyse the process. But this is highly speculative.

## 6. Conclusions

Knowing how common concepts work and how they are appraised and evaluated may be a key step on the road to learning how best to improve the stock of common concepts people use in everyday life, as conceptual engineers aim to do. Techniques could be developed to increase or decrease the feeling that a concept is dependable. Other techniques could help people to greater awareness of when their understanding of a concept is weak or unanalysed. I end by noting that these potential methods would be of great importance for higher education, in the development of the student's critical awareness of the strength and weakness of different concepts and the arguments that depend on them. Teaching the students to also recognize and debate the issue of when a

concept is “objective” and externally determined, when it is a social construction, open to cultural variation and possible manipulation, and when it is a matter for opinion, will be an important method for developing the skills of critical thinking.

**References**


Ahn, W., Kalish, C., Gelman, S. A., Medin, D. L., Luhmann, C., Atran, S., Coley, J. D., Shafto, P. (2001). Why essences are essential in the psychology of concepts. *Cognition, 82*, 59-69.

Barsalou, L. W. (1989). Intraconcept similarity and its implications for interconcept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 76–121). Cambridge University Press. https://doi.org/10.1017/CBO

Barsalou, L. W., & Hale, C. R. (1993). Components of conceptual representation: from feature lists to recursive frames. In I.van Mechelen, J. A. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and Concepts: Theoretical Views and Inductive Data Analysis* (pp. 97-144). London: Academic Press.

BBC News (26 October 2017) Card game bridge not a sport, European court rules. https://www.bbc.co.uk/news/uk-england-beds-bucks-herts-41762824

Browne, C. (2022). Duelling Diagnoses. *Association for Psychological Science, Observer*, June 30, 2022. https://www.psychologicalscience.org/observer/dueling-diagnoses. Accessed 24 March 2023.

Burge, T. (1979). Individualism and the Mental. *Midwest Studies in Philosophy* 4(1), 73-122.

Candlish, S., & Wrisley, G. (2019) "Private Language", *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2019/entries/private-language/

Cappelen H. (2018). *Fixing Language: Conceptual Engineering and the Limits of Revision*. Oxford: Oxford University Press.

Carnap, R. (1947). *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press.

Cronin, M. (2014) Amateurs and professionals (Chapter 3). *Sport: A Very Short Introduction.* Oxford: Oxford Academic. https://doi.org/10.1093/actrade/9780199688340.003.0004

De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M.J., Voorspoels, W., & Storms, G. (2008) Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods* 40(4), 1030–1048.

Dennis, I., Hampton, J. A., & Lea, S. E. G. (1973). New Problem in Concept Formation. *Nature, 243*, 101-102.

Djalal, F. M., Hampton, J. A., Storms, G., & Heyman, T. (2018). Relationship between Extensions and Intensions in Categorization: A Match Made in Heaven? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44,* 655-666.

Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong*. Cambridge, MA: MIT Press.

Fodor, J. A., & Lepore, E. (2002). *The Compositionality Papers*. Oxford: Oxford University Press.

Gelman, S. A. (2004). Psychological essentialism in children. Trends in Cognitive Sciences*.* 8: 404-9. https://doi.org/10.1016/j.tics.2004.07.001

Hampton, J. A. (1979). Polymorphous Concepts in Semantic Memory. *Journal of Verbal Learning and Verbal Behavior, 18*, 441-461.

Hampton, J. A. (1982). A demonstration of intransitivity in natural categories. *Cognition, 12*, 151-164.

Hampton, J. A. (1987). Inheritance of Attributes in Natural Concept Conjunctions. *Memory & Cognition, 15,* 55-71.

Hampton, J. A. (1988a). Disjunction of Natural Concepts. *Memory & Cognition, 16*, 579-591.

Hampton, J. A. (1988b). Overextension of Conjunctive Concepts: Evidence for a Unitary Model of Concept Typicality and Class Inclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 12-32.

Hampton, J.A. (1991). The combination of prototype concepts. *In* P. Schwanenflugel (Ed.) *The psychology of word meanings.* Hillsdale, NJ: Erlbaum.

Hampton, J. A. (1995). Testing Prototype Theory of Concepts. *Journal of Memory and Language, 34,* 686-708.

Hampton, J. A. (1997). Conceptual combination: conjunction and negation of natural concepts. *Memory & Cognition, 25*, 888-909.

Hampton, J. A., & Jönsson, M. L (2012). Typicality and compositionality: The logic of combining vague concepts. In M.Werning, W.Hintzen, and E.Machery (Eds.), *Oxford Handbook of Compositionality.* Oxford: Oxford University Press, pp 385-402.

Hampton, J. A., & Passanisi, A. (2016). When intensions don't map onto extensions: Individual differences in conceptualization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 505-523.

Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology, 39*(1), 113–127. https://doi.org/10.1348/014466600164363

His Majesty's Revenue and Customs HMRC (updated 16 February, 2023) VFOOD6260 - *Excepted items: Confectionery: The bounds of confectionery, sweets, chocolates, chocolate biscuits, cakes and biscuits: The borderline between cakes and biscuits.* https://www.gov.uk/hmrc-internal-manuals/vat-food/vfood6260.

Lawson, R. (2006). The science of cycology: failures to understand how everyday objects work. Memory & cognition, 34(8), 1667-1675. https://doi.org/10.7551/mitpress/1602.001.0001

Malt, B. (1994). Water is not H2O. *Cognitive Psychology, 27,* 41-70.

McLeod J, & Sundet, R. (2022). Psychotherapy as making. *Frontiers in Psychology*. Nov 17;13:1048665. doi:10.3389/fpsyg.2022.1048665

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). Cambridge University Press. https://doi.org/10.1017/CBO9780511529863.009

Murphy, G.L. (2002). *The Big Book of Concepts.* Cambridge: MIT Press. https://doi.org/10.7551/mitpress/1602.001.0001

Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition, 11*, 35-58.

Patalano, A. L., & Ross, B. H. (2007). The role of category coherence in experience-based prediction. *Psychonomic Bulletin & Review* **14**, 629–634. https://doi.org/10.3758/BF03196812

Putnam, Hillary (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science 7*, 131-193.

Quine, W. V. O. (1970). *Philosophy of Logic.* Cambridge MA: Harvard University Press.

Rehder, B., & Kim, S. (2010). Causal status and coherence in causal-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1171–1206. https://doi.org/10.1037/a0019765

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21-59). Cambridge: Cambridge University Press.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*(3), 192–233. https://doi.org/10.1037/0096-3445.104.3.192

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.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: an illusion of explanatory depth. *Cognitive science*, *26*(5), 521–562. https://doi.org/10.1207/s15516709cog2605_1

Serpico D. The Cyclical Return of the IQ Controversy: Revisiting the Lessons of the Resolution on Genetics, Race and Intelligence. J Hist Biol. 2021 Jun;54(2):199-228. https://doi.org/ 10.1007/s10739-021-09637-6.

Smits, T., Storms, G., Rosseel, Y., & De Boeck, P. (2002). Fruits and vegetables categorized: An application of the generalized context model. *Psychonomic Bulletin & Review, 9*, 836-844. https://doi.org/10.3758/BF03196343.

Thorne, S. R., Quilty-Dunn, J., Smortchkova, J., Shea, N., & Hampton, J. A. (2021). Concept Appraisal. *Cognitive Science, 45(5),* 1-35. https://doi.org/10.1111/cogs.12978

Urmson, J.O. (1970). Polymorphous Concepts. In: Wood, O.P., Pitcher, G. (eds) *Ryle. Modern Studies in Philosophy*. Palgrave, London. https://doi.org/10.1007/978-1-349-15418-0_11

Welty, C., & Smith, B., Eds. (2001). *Formal Ontology and Information Systems*. New York: ACM Press.

Wierzbicka, A. (1984). Apples are not a "kind of fruit": the semantics of human categorization. *American Ethnologist, 11,* 313-328. https://doi.org/10.1525/ae.1984.11.2.02a00060

Wittgenstein, L. (1953). Philosophical investigations. New York: MacMillan.

Xiao Y., Baes, N., Vylomova, E., & Haslam, N. (2023). Have the concepts of 'anxiety' and 'depression' been normalized or pathologized? A corpus study of historical semantic change. *PLOS ONE 18(6):* e0288027. https://doi.org/10.1371/journal.pone.0288027