On categorial membership

Michael Freund
IHPST
University of Paris Sorbonne
13 rue du Four 75006 Paris, France
micfreund@gmail.com

Abstract

We investigate the family of concepts that an agent comes to know through a set of defining features, and examine the role played by these features in the process of categorization. When categorial membership is measured through a weak order relation translating the fact that a concept may apply more to an object than to another, the passage from the features membership functions to this global ordering poses a problem analogous to vote aggregation in social choice theory. This similarity leads to an original solution that is particularly well-adapted to the framework of cognitive psychology. The resulting membership order extends to compound concepts, and provides a good description of the guppy paradoxe and the conjunction effect.

Keywords categorization, concept, extension, categorial membership, guppy effect, conjunction effect, word meaning, social choice, preference aggregation.

1 Introduction

An important domain of categorization theory is that of categorial membership, which analyzes the link between concepts and objects and the way the former apply to the latter. The original Fregean view (Frege, 1879), which reduced categorial membership to an all-or-nothing matter, was thereafter enlarged to that of a graded notion with intermediate states: as pointed out

by Eleanor Rosch, membership became a 'matter of degrees' (Rosch, 1975). In this perspective, the fuzzy logic of Zadeh (Zadeh, 1965) seemed to provide the adequate tool for the treatment of categorial membership, replacing the notion of membership degrees by that of membership functions taking their values in the unit interval. Some drawbacks and counterintuitive results pointed out by Osherson and Smith (Osherson and Smith, 1981) and Kamp and Partee (Kamp and Partee, 1995) drove the researchers to revise the initial fuzzy model, and to look for alternative fuzzy logics - see (Dubois et al., 2005) or (Lee, 2003) for an overview on the most recent work in this area.

Our approach differs from the preceding ones in that we consider categorial membership as a *qualitative* rather than a *quantitative* notion. As we argued in (Freund, 2008) and (Freund, 2009), the strength with which a concept applies to an object is best accounted for by a (pre)order relation among the objects of the universe: a given agent may be unable to determine to which degree a machine-gun may be considered as a *weapon of mass destruction* (WMD), while judging that it falls *more* under the concept of WMD than a cross-bow. From our point of view, categorial membership has therefore to be accounted for by means of membership orders.

In this paper, we shall restrict our attention to the family of concepts that come to an agent's knowledge through a set of *descriptive* or *defining features*. This family is far from covering the class of all concepts, but it is still large enough to deserve a treatment of its own. We shall try to understand how, for a concept of this type, categorial membership may be inherited from its features. This situation is analogous to that encountered in social choice theory or in multi-criteria decision theory, where a general preference order has to be built by aggregating individual preferences. In this perspective, an original solution will be proposed, leading to a construction that conforms with the intuition and is free from several drawbacks encountered in previous attempts. Its compositional properties put a new light on the so-called 'guppy effect' which, together with the 'conjunction fallacy', is at the center of most of the recent work in concept composition.

This paper is organized as follows: in Section 2, we recall some basic facts concerning membership orders, their use and utility. The framework of our study is specified in Section 3, where we discuss the difference between concepts and features, introducing the notion of *featured concepts*. In Section 4, the central part of this paper, we show how to build a membership order for concepts defined by their features. Two solutions are proposed, and their

properties compared. We also show how to build from these (partial) orders a membership distance function, which can be thereafter used to build a membership degree function. The case of compound concepts is examined in Section 5: there, we define the membership order that can be attached to a composed concept obtained by determination or modification of a principal concept; we also evoke some paradoxes of the classical theory and show that they can receive a correct treatment in the framework of qualitative membership orders. In Section 6, we show that the theory can be enlarged to a more general family of concepts, covering the class of concepts that can be defined through the elementary grammar of a dictionary. Section 7 is a conclusion.

The proof of the Observations is given in the Appendix.

2 Membership orders

In general, the human mind has no tool at its disposal to directly evaluate the categorial membership of an object: even though, in the extremal cases, we may be able to decide that a given object is or is not a member of the concerned category, we do not know in the intermediate states how to quantify its partial membership: a conventional bomb is definitely not a *weapon-of-mass-destruction* (WMD), and the same is true for a machine gun, yet, we are unable to attribute a precise degree of WMD-membership to any of these items.

As a matter of fact, the only thing the human mind is capable of concerning membership evaluation is to *compare* two objects and decide which one, if any, falls 'more' under the concerned concept. Thus, the concept *to-be-a-weapon-of-mass-destruction* will be generally considered as applying *more* to a machine-gun than to an arquebus, and *less* to a spear than to an arquebus. Clearly, this judgement shows the existence of a basic ordering induced by the concept *to-be-a-weapon-of-mass-destruction* in the universe of discourse. This ordering is by no means a consequence of degree assignment that the agent has set *a-priori* on the objects at his disposal, although such an assignment becomes possible once the agent has gathered a collection of specimen and proceeded to compare them to each other. For instance the agent may decide to rank the weapons she knows into a non-decreasing sequence that reflects their relative ability to be considered as a WMD. A sequence like *bludgeon* $\leq$ *sword* $\leq$ *spear* $\leq$ *crossbow* $\leq$ *arquebus* $\leq$ *gun* $\leq$ *machine-gun* $\leq$ *flamethrower* $\leq$ *conventional bomb* $\leq$ *scud* $\leq$ *atomic bomb*

could thus be established, rendering possible the attribution of an explicit or implicit membership degree to each of the concerned items. If the agent understands this sequence as a strictly increasing one, she will consider an *atomic bomb* as being 100% a WMD, a *scud* as 90%, a *conventional bomb* as 80% and so on. If she considers that the six first items of the sequence are equivalent, being totally deprived of anything that evokes a WMD, the agent will rank the *scud* as a WMD at degree 80%, the *conventional bomb* at 60%, the *flamethrower* at 40% and the *machine-gun* at 20%, while *bludgeon*, *sword, spear, crossbow, arquebus* and *gun* are by no means to be considered as WMD. The point is that, in any case, these WMD-degrees appear as a consequence of a pre-recognized order among a given set of weapons: they do not stand at the origin of it.

Order relations therefore appear to provide the most adequate model to account for categorial membership as perceived by a cognitive agent. Appealing systematically to relations of this type whenever it is possible avoids the drawbacks that may result from the application of more sophisticated theories. In some cases, order relations may be insufficient to fully treat categorial membership - this will be for instance the case for *fuzzy concepts*, or for vague concepts of a continuous type: for these concepts, interesting theories have been developed in different domains - fuzzy set theory (Zadeh, 1982), geometrical spaces (Gärdenfors, 2000), and even quantum mechanics (Franco, 2009) and (Aerts, 2009). But for the specific class of concepts that will be studied in this paper, the family of order relations is sufficiently wide to adequately model the problem of categorial membership.

We need at this stage to specify what we mean by *membership order*: the fact that a concept α may apply *more* (or *better*) to an item x than to an item y is best translated by a *preorder* relation, that is a reflexive and transitive relation on the set of objects $\mathcal{O}$ of the universe. This relation will be denoted by $\preceq_\alpha$. We shall therefore consider that the expression '$x \preceq_\alpha y$' translates the sentence 'the concept α applies at least as much to the object y as to the object x', or, equivalently 'x falls no more under the concept α than y'. We shall denote by $\prec_\alpha$ the corresponding strict partial order, that is the relation '$x \preceq_\alpha y$ and not $y \preceq_\alpha x$'. We can then translate the expression '$x \prec_\alpha y$' by the sentence 'the concept α applies more to y than to x', or 'x falls less under α than y'. The equivalence relation induced by $\preceq_\alpha$ will be denoted by $\sim_\alpha$: we have therefore $x \sim_\alpha y$ if and only if $x \preceq_\alpha y$ and $y \preceq_\alpha x$.

In the general case the relation $\preceq_\alpha$ will not be supposed to be total: in this we depart from most of the existing theories. Indeed, we consider that

there is no reason to *a priori* eliminate the case where the membership of two objects is incomparable - which may be for instance the case when the *birdhood* of a tortoise is compared with that of a bat .

A construction of a concept membership order has been already presented in (Freund, 2008) and (Freund, 2009) in the particular case of *definable concepts*, that is concepts that could be inductively defined from sharp concepts like in a rudimentary dictionary. In this paper, we propose a different construction, simpler than the original one, and applicable to a larger class of concepts.

2.1 Finite membership orders

A (pre)order $\preceq$ will be said to be *finite* if it has only finitely many classes for the associated equivalence relation. For such orders, any increasing chain is of bounded length. Such is for instance the case for membership orders associated with fuzzy concepts that admit a *discrete* membership degree function. As a limit case, the simplest example is provided by *sharp concepts*, for which the membership order has only two classes: one constituted by the objects that fall under the concept, the other one by the objects that do not fall under it. We shall justify in the next section the choice of these orders to model concept membership. From a computational point of view, finite membership orders are of particular interest because, even though they are not total orders, they give rise to a natural notion of *membership distance* and of *membership degree*, as we show now:

Let α be a concept with finite associated membership order $\preceq_\alpha$, and denote by $Ext\,\alpha$ (the *extension* of α) the set of all $\prec_\alpha$-maximal objects [1]. Given an object x, we first define its *α-membership distance* $\mu_\alpha(x)$ as the maximal length of a chain $x \prec_\alpha x_1 \prec_\alpha x_2 \prec_\alpha \ldots \prec_\alpha x_n$ starting from x. Such a chain ends up with an element $x_n \in Ext\,\alpha$. It measures how far, for a given agent, an object stands from a category. We have clearly $\mu_\alpha(x) > \mu_\alpha(y)$ whenever $x \prec_\alpha y$, but the converse is not necessarily true. Note that $\mu_\alpha(x) = \mu_\alpha(y)$ whenever $x \sim_\alpha y$, and that $x \in Ext\,\alpha$ if and only if $\mu_\alpha(x) = 0$.

The membership distance yields a membership degree, as we saw in WMD example. A *membership degree function* δ_α can be indeed defined by setting,

[1] We will show later that the extension of a concept exactly corresponds to the set of all objects that fall under this concept.

for every object x, $\delta_\alpha(x) = 1 - \frac{\mu_\alpha(x)}{N_\alpha}$, where N_α is a constant that can be chosen as the maximal length of an $\prec_\alpha$-chain. Again, this length may vary between two agents, leading to different degrees δ_α and δ'_α even in the case where the agents agree on the membership order $\preceq_\alpha$. The degrees are then related through the equality $1 - \delta_\alpha = k(1 - \delta'_\alpha)$, with $k = \frac{N'_\alpha}{N_\alpha}$. In the WMD examples for instance, we had $k = 1/2$.

Note that $\delta_\alpha(x) = 1$ if and only if $x \in Ext\,\alpha$, that $\delta_\alpha(x) = 0$ if and only if x is $\prec_\alpha$-minimal, and that $\delta_\alpha(x) < \delta_\alpha(y)$ whenever $x \prec_\alpha y$.

3 Featured concepts

Following the *attributional view* which gave rise to the so-called *binary model* (Smith et al., 1974), (Smith and Medin, 1981), and (Osherson and Smith, 1982), *class membership* relative to a concept is accounted for by a set of *defining features*, while all questions concerning *typicality* are taken care of through a (different set) of *characteristic features*. It is only the former set that will retain our attention in the present paper since our aim is actually to provide a framework for categorial membership. (For a general treatment of typicality, see (Freund, 2009)). Numerous counter examples however showed that, contrary to this theory, arbitrary concepts cannot be generally reduced to a list of constituents. Fodor (Fodor, 1998) argues for instance that here exists practically no examples for successful definition around. Our position stands between these two extremes: although it is clear that not all concepts can be defined or described through a set of features, this remains still the case for some important subclasses, like those of *nominal*, *natural kind* or *scientific* concepts. They constitute a family large enough to deserve a study of its own.

In an agent's mind, the defining features attached to a concept are those by which the concept may be fully seized and unsderstood: they are *individually necessary and collectively sufficient* to specify membership in a category (Rosch and Mervis, 1975). In particular, an object falls under a concept if and only if it falls under all its defining features. It is worth noticing that this supposes the existence of an item on which all the defining features of this concept apply simultaneously.

The attributional view links the categorization process relative to a concept with categorization relative to its defining features. For example, if for a given agent the defining feature set associated with the concept *to-be-a-bird*

consists of the features *to-be-warm-blooded*, *to-have-a-beak*, *to-have-feathers* and *to-have-wings*, the *birdhood* of a given animal is to be analyzed through its membership relative to theses features. Similarly, to quote an example of Putnam (Putnam, 1975), the *meaning* of the term *tiger* is to be searched by reference to the terms *yellow feline, fierce, black stripes,* and *jungle*: the defining features associated with the concept *to-be-a-tiger* consist of the properties *to-be-a-yellow-feline, to-have-black-stripes, to-live-in-the-jungle* and *to-be-fierce*.

Of course, the defining features of a concept may vary from an agent to another. It may even be the case that, for a given agent, a concept has an attached set of defining features, while being deprived of such a set for another agent. For this reason, our work is o be analyzed as an attempt to model the categorization process of a single particular given agent at a well defined moment.

The concepts we are interested in are, in a first place, those for which, similarly to the attributional view, membership is measured through a set of features. At a second stage we shall consider concepts whose definitions are of a more elaborated type. This will encompass a more general family, large enough to deserve a treatment of its own, even though it dos not cover the whole set of concepts that are part of an agent's knowledge.

3.1 Concepts and features

The terms *concept* and *feature* cover different notions. Formally, in English language, concepts are most often introduced through the auxiliary *to-be* followed by a noun: *to-be-a-bird, to-be-a-vector-space, to-be-a-democracy*. Features may be presented through a verb (*to-fly*), the auxiliary *to-have* followed by a noun (*to-have-a-beak)*, or the auxiliary *to-be* followed by an adjective (*to-be-tall*). While concepts appear as unary predicates, this condition is no more necessary for features, which may take arbitrary forms. On the ground level, we know that features, like concepts, apply to the objects at hand, but, contrary to concepts, they borrow part of their significance from the concept they are attached to. Properties like *to-be-tall*, *to-be-rich*, or *to-be-red* take their full meaning only in a given context, that is when qualifying a well-defined entity. Even simple verbal forms like *to-fly*, *to-run*, *to-live-in-water*, *to-be-made-of-metal* need a principal referent concept to fully seize the strength with which they apply to different items. Thus, the concept to which the feature applies may be seen itself as a contextual determination of

this feature. To summarize, we consider that the meaning of a feature depends on the context in which this feature is used, contrary to the meaning of a concept, which exits by itself.

It does not seem at this stage that any formalism can fully account for the difference between features and concepts. It is true that in Description Logics, different treatments are applied to one and two-places predicates: binary predicates characterize indeed the *roles* of the language, which are used to express relationship between the concepts (Nardi and Brachman, 2003). Thus, *to-be-a-tree* will be a concept, expressible by a single symbol A, but *to-have-green-leaves* is a 'role', expressed by a formula of the type '$\forall$ hasLeaves.Green'. However, no difference is made between unary predicates that translate a notion of concept and those that translate a notion of feature.

Note finally that, when considering a concept α and a feature f of α, the strength with which f applies to an object may be itself related to the categorial membership of an auxiliary concept β. For instance, in order to evaluate to which degree an item x may be considered as *having-wings*, we must be able to determine what exactly covers the concept *to-be-a-wing*, and which objects fall under it (for a discussion on the difference between *knowing what an X is and knowing what it is to have an X*, see (Fodor, 1998)). This observation shows that circularity cannot be avoided when trying to build a general mathematical model for categorial membership. However, such is not the purpose of this work: in this paper, we only try to model the way categorial membership of concepts can be deduced from the knowledge of their defining features, without questioning the nature and the sources of this knowledge.

3.2 Applicability functions and applicability orders for concept features

In a given ontology, the strength with which a feature f applies to an object is usually measured through an *applicability degree function* δ_f with values in the unit interval, satisfying $\delta_f(x) = 1$ if f 'fully' applies to x, and $\delta_f(x) = 0$ if x is totally deprived of f. Concerning the range of δ_f, it is important to note that, most often, it can be circumscribed to a *finite* subset of $[0, 1]$: this is clearly true for fuzzy features like *to-be-tall*, *to-be-rich* or *to-be-warm*, since the measure of their applicability is always approximative (to an inch, a cent or a degree). In Britain coast, for instance, the set of displayed water

temperatures t in July ranges from 15 degrees Celsius to 25 degrees Celsius, thus covering 10 possible values. The function associated with the concept *cold* taken in this context may be given by $\delta_f^\alpha = -t/10 + 2.5$.

The finiteness of the range of δ_f is even more obvious in the general process of categorization: in the context of a given concept α, ranking the objects relatively to a feature yields only a small number of non equivalent classes. To determine, for instance, to which extent a flower may be considered as a *poppy*, one may, among others, evaluate its redness. Comparison with other real or fictitious items shows only a finite number of non equivalent reds that stand between the color of that particular flower and that of an ideal poppy. Thus, in the context of a *poppy*, there exists only a small number of possible degrees of *redness*.

For this reason, we shall restrict our attention to the family of concepts whose defining features can be weighed on a finite scale. We shall call these concepts *featured concepts*. Given such a concept α, the way any of its defining features f applies to an item can be therefore accounted for by an applicability function δ_f^α that takes only a finite number of values. Equivalently, we may say that f, as an α-feature, generates on the set of objects at hand a *finite and total* applicability (pre)order $\preceq_f^\alpha$, defined by: $x \preceq_f^\alpha y$ iff $\delta_f^\alpha(x) \leq \delta_f^\alpha(y)$. Given an object x totally deprived of f and an object y to which f fully applies, there exists only a finite number of intermediate states.

Putting together what precedes, we are now ready to introduce the notion of *featured concepts*:

Definition 1 *A featured concept α is a concept for which there exists a set of defining features $\Delta(\alpha)$ satisfying the two properties :*

1. *for every feature f of $\Delta(\alpha)$, the corresponding applicability function δ_f^α takes a finite number of values.*

2. *There exists at least an item z such that $\delta_f^\alpha(z) = 1$ for all $f \in \Delta(\alpha)$.*

In order to lighten the notations, we shall omit the superscript, and simply write $\preceq_f$ and δ_f for $\preceq_f^\alpha$ and δ_f^α. However, we have to keep in mind that applicability orders or functions are always defined in the frame of a given concept.

We shall write $x \sim_f y$ for ($x \preceq_f y$ and $y \preceq_f x$).

4 Categorial membership and preference aggregation

We suppose given a featured concept α together with its set of defining features $\Delta(\alpha) = \{f_1, f_2, \ldots, f_k\}$. In the perspective of the *attributional view*, the categorization process relative to α is tightly linked with the categorization process relative to its defining features. This means that the knowledge of the applicability orders induced by $(f_1, f_2, \ldots, f_k)$ is sufficient to determine the membership order associated with α. In this section, we shall try to understand how one completes the passage from the $\preceq_{f_1}, \preceq_{f_2}, \ldots, \preceq_{f_k}$ to the target order $\preceq_\alpha$.

It is interesting to observe that the problem of determining $\preceq_\alpha$ from the $\preceq_{f_i}$'s is closely related with that encountered in *social choice* or in *multicriteria decision theory*: there indeed, one tries to aggregate individual preferences concerning a certain number of items into a general preference that would best approach the individual ones. Let us give two examples:

Example 1 *Suppose you want to buy an apartment in a given town. You are offered a list of several possibilities. You will make up your mind after evaluating the apartments of the list relatively to a certain amount of criteria $(f_1, f_2, \ldots, f_k)$ - the price, the surface, the district, the condition of the place, the view and the existence of some items : elevator, parking possibilities, closeness of a metro or a bus station. Once you have compared the plus and minus of each of the proposed apartments in the light of these criteria, you can rank the elements of the list and eventually choose the best one. Each criterium f_i induced an order $\preceq_{f_i}$ on your list, and a final order was obtained by aggregating together these orders.*

Example 2 *Electoral systems are confronted with the problem of aggregating individual votes in the situation where several candidates have to be ranked by a set of voters $(f_1, f_2, \ldots, f_k)$. Each voter f_i proposes a ranking $\preceq_{f_i}$ among the set of candidates, and the task of the electoral system is to aggregate together these orders so that the resulting ranking best reflects the individual preferences of the voters.*

Since the original work of Borda (1781) and Condorcet (1785), a considerable amount of research has been done on the problem of aggregating individual preferences into a ranked order. In 1953, Arrow's impossibility theorem (Arrow, 1953) showed that no solution coud be expected that would satisfy the

three axioms of *non-dictatorship* (no voter can impose its preferences regardless to other voters), *unanimity* (if all voters prefer y to x, then y should be preferred to x in the final order) and *independence* (the final preference between two items does not depend on the rank of the other items). In the framework of categorial membership, things, though, are different. First, as mentioned before, the relation to be built is not required to be total. Second, the Arrow axioms need not be all satisfied: it is clear that a property like *unanimity* should hold in the framework of concept membership - if all the features of a concept apply more to an object y than to an object x, then it should be the case that the concept itself applies more to y than to x - but the two other Arrow properties are of less importance.

To fix the ideas, let us quickly examine two solutions that can be proposed at a basic level.

The Pareto rule

The simplest way to build the relation $\preceq_\alpha$ from the relations $\preceq_{f_i}$ is to take their intersection, setting $\preceq_\alpha = \bigcap_{f_i \in \Delta(\alpha)} (\preceq_{f_i})$. This amounts to considering that the concept α applies no more to x than to y if such is the case for each of its defining features. Note that this construction satisfies all Arrow's properties. However, in this very skeptical approach, any feature f_i of α can be considered as a *weak dictator* : if for some i one has $x \prec_{f_i} y$, then it cannot be the case that $y \preceq_\alpha x$, no matter how x and y stand relatively to the other features.

The Borda method

In this method, the membership order induced by α rests on the sum of the applicability degrees relative to each feature. That is, we have $x \leq_\alpha y$ iff $\sum \delta_{f_i}(x) \leq \sum \delta_{f_i}(y)$. The axioms of *non-dictatorship* and *unanimity* are satisfied, but this is not the case for that of *independence*: the preference between x and y depends on the associated degrees $\delta_{f_i}(x)$ and $\delta_{f_i}(y)$, and these are attributed after ranking all the elements of $\mathcal{O}$.

Coming back to the general case, we note that although there exist formal similarities between the classical problems of social choice theory and that of categorial membership for featured concepts, there exists an important difference: contrary to voters, whose votes have all the same weight, the defining features of a concept are usually not considered as equivalent. Some features are more salient than others, and the set $\Delta(\alpha)$ is usually presented with a *salience* relation between its elements. For instance, a particular agent that has associated with the concept *to-be-a-bird* the defining set {*to-be-a-*

vertebrate, to-be-oviparous, to-have-feathers, to-have-a-beak, to-have-wings$\}$ may consider that *having wing*s is a more important feature for birdhood than *having a beak*. From her point of view, a bat will be endowed with more birdhood than a tortoise.

We have therefore to take into account the relative salience of the elements of $\Delta(\alpha)$. When this salience is quantifiable, that is, when it can be translated by a natural number corresponding to a degree or to a rank of importance, this can be done through a simple perequation, the voice of a voter of rank i weighing i times that of an 'ordinary' voter. For instance, a weight $w_\alpha(f_i)$ being attributed to each feature f_i, a natural extension of the Borda method, would yield $x \leq_\alpha y$ iff $\sum_i w_\alpha(f_i)\delta_{f_i}(x) \leq \sum_i w_\alpha(f_i)\delta_{f_i}(y)$ (see (Hampton, 1993) or (Hampton, 1995)).

However, there is no reason why the defining features of a concept should be attributed such a numerical rank of importance. Again, an agent may be quite able to compare the relative salience of two features of a concept without being able to associate a *degree* to any of these saliences. In the most general case, the salience of a defining feature should be considered as a *qualitative* notion. We shall therefore simply consider that the salience relation among the features of α is translated by an arbitrary strict partial order on the set $\Delta(\alpha)$, which will be denoted by $>_s$.

Coming back to our analogy, we must now build a voting procedure that takes into account the hierarchy that may exist inside a group of voters. We propose two solutions to this problem.

4.1 The lexicographic order

The idea is that the voice of a voter can be overruled by any of its hierarchical superior: a candidate x will be considered as at most as good as a candidate y if, for each voter that prefers x to y, there exists a more important voter that prefers y to x.

Formally we set

$x \preceq_\alpha y$ iff for each defining feature f_i such that $y \prec_{f_i} x$, there exists a defining feature f_j, $f_j >_s f_i$, such that $x \prec_{f_j} y$.

It is immediate that this relation is reflexive. The connectedness of the $\preceq_{f_i}$'s implies that it is also *transitive* and therefore constitutes a (partial) preorder:

Observation 1 *The relation $\preceq_\alpha$ is transitive.*[2]

The strict partial order induced by $\preceq_\alpha$ reads

$$x \prec_\alpha y \text{ if and only if } x \preceq_\alpha y \text{ and for some defining feature } f_i, \text{ it holds } x \prec_{f_i} y.$$

Note that $\preceq_\alpha$ satisfies the *unanimity* and the *independence* rules, but not the *non-dictatorship* one: if a feature f dominates all other ones, one has $x \preceq_\alpha y$ whenever $x \preceq_g y$.

The equivalence relation $\sim_\alpha$ induced by $\preceq_\alpha$ is easily seen to be the intersection of the $\sim_f$, for $f \in \Delta(\alpha)$:

Observation 2 *One has $x \sim_\alpha y$ if and only if $x \sim_f y$ for all features f of $\Delta(\alpha)$.*

Example 3 *Let α be the concept* to-be-a-bird*, and suppose that, from the point of view of an agent, its defining feature set, in the context of* being-an-animal*, is given by $\Delta_\alpha =$* {to-have-two-legs, to-lay-eggs, to-have-a-beak, to-have-wings}*, all of these features being considered as two-valued functions for the agent. Suppose also that the salience of these features is as follows:* to-have-a-beak $>_s$ to-lay-eggs $>_s$ to-have-two-legs*, and* to-have wings $>_s$ to-lay-eggs.

Let s, m, t, b and d respectively stand for a sparrow, a mouse, a tortoise, a bat and a dragonfly, and let us compare their relative birdhood. In order to determine the induced membership order, we first build the following array:

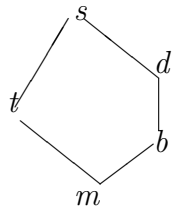
	$two-legs$	$lay-eggs$	$beak$	$wings$
sparrow	$\star$	$\star$	$\star$	$\star$
mouse				
tortoise		$\star$	$\star$	
bat	$\star$			$\star$
dragonfly		$\star$		$\star$

We readily check that $d \prec_\alpha s$, $m \prec_\alpha t$, and $m \prec_\alpha b$. Note that we have $b \preceq_\alpha d$, since the concept to-have-two-legs *under which the bat falls, contrary to the dragonfly, is dominated by the concept* to-lay-eggs *that applies to the*

[2]We recall that the proof of the observations is given in the appendix.

dragonfly and not to the bat. On the other hand, we do not have $d \preceq_\alpha b$*, as nothing compensates the fact that the dragonfly lays eggs and the bat does not. This yields* $b \prec_\alpha d$*. We also remark that the tortoise and the bat are incomparable, that is, we have neither* $b \preceq_\alpha t$*, nor* $t \preceq_\alpha b$*.*

The strict α*-membership order induced by the agent on these five elements is thus given by the following Hasse diagram:*



Coming back to the construction of $\preceq_\alpha$ let us examine how it behaves in the two following limit cases:

- Suppose that the salience order on $\Delta(\alpha)$ is empty. We have then $x \preceq_\alpha y$ iff $x \preceq_f y$ for all elements of $\Delta(\alpha)$. Thus, y falls at least as much as x under α, if and only if every single defining feature of α applies at least as much to y as to x. We retrieve here $\preceq_\alpha$ as an intersection: $\preceq_\alpha = \bigcap_i \preceq_{f_i}$.

  Note that in the case of an empty salience order, the Borda-Hampton method evoked above yields $x \leq_\alpha y$ iff $\sum_{f\in\Delta(\alpha)} \delta_f(x) \leq \sum_{f\in\Delta(\alpha)} \delta_f(y)$. This may lead to counterintuitive or undesirable effects: if for instance α has three defining features f, g, and h, whose appliance degrees on two items x and y are $\delta_f(x) = \delta_g(x) = 0.7, \delta_h(x) = 1$; $\delta_f(y) = \delta_g(y) = 1, \delta_h(y) = 0$, then, one gets $y <_\alpha x$ although two out of the three defining features of α apply more to y than to x.

- Suppose now that the salience order on $\Delta(\alpha)$ is total. Let x and y be two arbitrary items. Then, either we have $x \sim_\alpha y$, or, by Observation 2, there exists a defining feature f_i for which it is not the case that $x \sim_{f_i} y$. We can suppose that f_i is of maximal salience for this property. We have then $x \prec_\alpha y$ or $y \prec_\alpha x$ depending whether $x \prec_{f_i} y$ or $y \prec_{f_i} x$. This means that the most salient voter that distinguishes between x and y dictates his opinion. Note that $\preceq_\alpha$ is then a *total* preorder.

The lexicographic order provides a satisfactory answer for the construction of a natural membership order for featured concepts. It can be built in the presence of a purely qualitative salience order, and appears to be free from the drawbacks encountered in classical solutions. One of its features however may appear to be undesirable, which is the preeminence it gives to *hierarchy*, independently of the number of voices that stand for or against a candidate. Suppose for instance that the features in $\Delta(\alpha)$ are of equivalent salience except for one, f_i, that dominates all the other ones. Let x and y be two objects such that $x \prec_{f_j} y$ for all $j \neq i$, but $y \prec_{f_i} x$. Then the hierarchical order will necessarily yield $y \prec_\alpha x$. The object x will be considered as falling more under α than y for the only reason that the dominant feature applies more to x than to y. The fact that f_i is more salient than any individual other feature f_j makes it a dictator, that has more weight than all the f's taken together.

Although the cases where $\Delta(\alpha)$ consists of one single salient feature opposed to a bunch of non salient defining properties are rare, it may be useful to consider a slight modification of the lexicographic order in order to also account for the number of voters that prefer candidate x to candidate y.

4.2 The proportional order

The idea is to prevent a same single individual from overruling several subordinate voters. For this purpose, we shall first consider the relation $\preceq'_\alpha$ defined by:

$x \preceq'_\alpha y$ if, for each sequence of distinct elements $f_1, f_2, \ldots f_n$ of $\Delta(\alpha)$ such that $y \prec_{f_i} x$, there exists a sequence of distinct elements $g_1, g_2, \ldots g_n$ of $\Delta(\alpha)$, $g_i >_s f_i$, such that $x \prec_{g_i} y$.

In this situation, each voter may see his decision overruled by a personally attached direct superior, but the voice of a single voter, be it the most important of all, cannot overrule more than one voice.

The relation $\preceq'_\alpha$ is clearly reflexive, but unlike $\preceq_\alpha$, it is generally not transitive, as can be seen from the following example:

Example 4 *Suppose that $\Delta(\alpha)$ consist of the three features f, g, h, with $h >_s f$ and $h >_s g$. Let x, y and z be three objects such that $z \prec_f x$, $x \sim_f y$, $y \sim_g z$, $z \prec_g x$, and $x \prec_h y \prec_h z$. Then one has $x \preceq'_\alpha y$ and $y \preceq'_\alpha z$ but not $x \preceq'_\alpha z$.*

Rather than the relation $\preceq'_\alpha$ itself, we have therefore to consider its transitive closure, which we shall denote by $\preceq^*_\alpha$. We will refer to it as to the *proportional order relation*. From its definition, we have thus $x \preceq^*_\alpha y$ iff there exists a sequence of k items $x_1 = x, x_2, \ldots, x_k = y$ such that $x_i \preceq'_\alpha x_{i+1}$.

The behaviour of $\preceq^*_\alpha$ in the case of an empty salience order on $\Delta(\alpha)$ is clearly identical to that of $\preceq_\alpha$: $\preceq^*_\alpha$ boils down to the intersection of all the membership orders induced by the defining features of α. In the second limit case, when the salience order is total, the proportional order differs from the hierarchical one: the most salient voter f_i that distinguishes between x and y no longer dictates his opinion, and the inequality $x \prec_{f_i} y$ is not sufficient to imply $x \preceq^*_\alpha y$.

In the general case, one easily sees that the proportional order is embedded in the lexicographic one: the inequality $x \preceq^*_\alpha y$ implies $x \preceq_\alpha y$. Similarly, the strict inequality $x \prec^*_\alpha y$ implies $x \prec_\alpha y$; the same is also true for the indifference relation: by Observation 2, $x \sim_\alpha y$ readily implies $x \sim^*_\alpha y$.

4.3 Some common properties of the lexicographic and the proportional orders

The first property enjoyed by the two α-membership orders is the common characterization of their maximal elements as the *extension* of the concept α:

Observation 3 *Define the* extension $Ext\,\alpha$ *of* α *as the set of all objects* z *such that* $\delta_f(z) = 1$ *for all* f *of* $\Delta(\alpha)$. *Then one has* $x \prec_\alpha z$ *and* $x \prec^*_\alpha z$ *for all* $x \notin Ext\,\alpha$ *and* $z \in Ext\,\alpha$.

Note that condition (2) of Definition 1 guarantees that $Ext\,\alpha$ is a non-empty set. Observation 3 asserts that the objects that fall under all defining features of α fall more under α than any other objects. They will be considered as forming the *category* associated with α. More precisely, we shall say that an object x 'falls' under the concept α if x is an element of its extension, that is if $x \in Ext\,\alpha$. This is equivalent to say that x is $\prec_\alpha$-maximal or $\prec^*_\alpha$-maximal among the objects that constitute the universe of discourse: there exists no object y such that $x \prec_\alpha y$ or $x \prec^*_\alpha y$. In this case, we may also say that x is an *instance* or an *exemplar* of α. This definition by means of maximal membership conforms with the intuition: an object x (fully) falls under a concept α if α cannot apply more to an object y than to x. It has

as a consequence that whatever salience order is set on $\Delta(\alpha)$, *an object falls under the featured concept α if and only if it falls under each of its defining features*. We retrieve here the classical characterization of a defining feature set, as a set of features that are individually necessary and jointly sufficient to ensure membership relative to α.

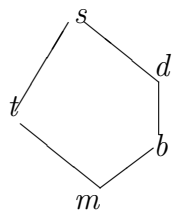
From the finiteness condition (1) we now deduce an important result:

Observation 4 *$\preceq_\alpha$ and $\preceq^*_\alpha$ are finite membership orders.*[3]

It is now possible to associate with every featured concept α a *membership degree* as was done in section 2.1. We denote by μ_α and μ^*_α the membership distances induced by $\preceq_\alpha$ and $\preceq^*_\alpha$, and by δ_α and δ^*_α the corresponding membership degrees.

We have then $x \in Ext\,\alpha$ if and only if $\mu_\alpha(x) = 0 = \mu^*_\alpha(x)$, that is if and only if $\delta_\alpha(x) = 1 = \delta^*_\alpha(x)$.

Example 5 *We take again Example 3 and the Hasse diagram giving the membership order induced by the concept to-be-a-bird on the set {sparrow, tortoise, bat, mouse, dragonfly}:*



Let us compute the different membership distances and degrees that an agent may attribute to the 5 items above: this agent may consider that $\mu_\alpha(t)$ = 1, not knowing of any oviparous animal a that has a beak and satisfies $t \prec_\alpha a \prec_\alpha s$. Similarly, she would set $\mu_\alpha(d) = 1$, since there exists no animal a' such that $d \prec_\alpha a' \prec_\alpha s$. Because the bat falls under two out of the five elements of Δ_α, one must have $\mu_\alpha(b) \leq 2$, and the inequality $b \prec_\alpha d \prec_\alpha s$ then yields $\mu_\alpha(b) = 2$. As for the mouse, it holds $m \prec_\alpha b \prec_\alpha d \prec_\alpha s$, but this is not a chain of maximal length. For instance, noting that macaques have two legs, the agent may consider the chain $m \prec_\alpha k \prec_\alpha b \prec_\alpha d \prec_\alpha s$, where k denotes a macaque. This shows that $\mu_\alpha(m) \geq 4$, and thus $\mu_\alpha(m) = 4$

[3]The original proof of this observation is due to Daniel Lehmann (Lehmann, 2010).

since, among the animals, only four simple features are sufficient to define a bird. This also shows that for our agent, in the context of animals, the longest possible birdhood-*chain has length* $N_\alpha = 4$ *Putting all this together, the agent will finally get* $\delta_\alpha(m) = 0, \delta_\alpha(k) = 1/4, \delta_\alpha(b) = 1/2, \delta_\alpha(t) = 3/4 = \delta_\alpha(d)$ *and* $\delta_\alpha(s) = 1$.

Let us now briefly examine what happens in the two limit cases of an empty or a total salient order :

- When the salience order is empty, we see that $\mu_\alpha(x) = \mu^*_\alpha(x) \leq \Sigma_f n_f(x)$, where $n_f(x)$ is the number of different values greater than $\delta_f(x)$ taken by the membership function δ_f.
- In case of a *total* salience order, we have seen that $\preceq_\alpha$ is a total order. From this, it follows that one has $\delta_\alpha(x) < \delta_\alpha(y)$ if and only if $x \prec_\alpha y$: the membership function exactly models in this case the membership order.

5 The case of compound concepts

Elementary concepts may aggregate through different ways to give birth to compound ones. The simplest one is the ordinary *conjunction*, which corresponds to a simple juxtaposition of terms. A more sophisticated connective was introduced in (Freund, 2008) to account for the *modification* or the *determination* of a concept by another one. The compound concepts we shall study are those obtained through conjunction or determination of simple concepts. Similarly to the simple case, membership relative to compound concepts can be described through a partial order among the objects of the universe. As we shall see, this order is closely related to the membership orders induced by each of the constituents of the compound concept. In this sense, categorial membership is a *compositional* notion that can be naturally extended in the case of concept conjunction as well as concept determination.

5.1 Categorial membership and concept conjunction

Taking the conjunction of two arbitrary concepts may lead to meaningless notions, like *(to-be-a-cloth)&(to-be-a-country)*. For this reason, we shall restrict ourselves to the case where the extensions of the components intersect.

Given two featured concepts α and β such that $Ext\,\alpha \cap Ext\,\beta \neq \emptyset$, the membership order on the conjunction $\alpha\&\beta$ is taken as the intersection of the membership orders of its components.

We have thus $x \preceq_{\alpha\&\beta} y$ iff $x \preceq_\alpha y$ and $x \preceq_\beta y$, and $x \preceq^\star_{\alpha\&\beta} y$ iff $x \preceq^\star_\alpha y$ and $x \preceq^\star_\beta y$.

Observe that in this construction, the feature sets of α and β are treated as if they had no common element. This comes from the fact that, as we emphasized previously, features borrow part of their significance from the concept they are attached to: a feature f considered as an element of $\Delta(\alpha)$ may receive a different treatment than the same feature considered as an element of $\Delta(\beta)$. This situation may be compared to that encountered in social choice in the case of a two-stages election, with first set of voters $\{f_1, f_2, \ldots, f_n\}$ and second set of voters $\{g_1, g_2, \ldots, g_k\}$: it may then happen that a same voter ranks the candidate x before the candidate y in the first election, and changes her mind at the second stage, setting y before x. In this case, x and y will be simply incomparable in the final result.

Similarly to what has been done is the case of elementary concepts, let us define the *extension* of the conjunction $\alpha\&\beta$ as the set of all $\prec_{\alpha\&\beta}$-maximal elements. The hypothesis made on $Ext\,\alpha$ and $Ext\,\beta$ implies that this set is also the set of all $\prec^\star_{\alpha\&\beta}$-maximal elements. We easily check that full membership is compositional in the sense that $Ext\,\alpha \cap Ext\,\beta = Ext\,(\alpha\&\beta)$.

The relations $\preceq_{\alpha\&\beta}$ and $\preceq^\star_{\alpha\&\beta}$ can be directly recovered by assigning to the concept $\alpha\&\beta$ a fictitious defining feature set equal to the disjoint union of $\Delta(\alpha)$ and $\Delta(\beta)$. More precisely, let $\widetilde{\Delta}(\alpha)$ be the set $\{(f_i, \alpha); f_i \in \Delta(\alpha)\}$ equipped with the salience order that makes $(f_i,\, \alpha)$ more salient than $(f_j,\, \alpha)$ if and only if f_i is more salient than f_j. Similarly, set $\widetilde{\Delta}(\beta) = \{(g_i,\, \beta); \delta \in \Delta(\beta)\}$ with the salience order that makes $(g_i,\, \beta)$ more salient than $(g_j,\, \beta)$ if and only if g_i is more salient than g_j. The structure of these sets emphasizes the fact that features depend of the concept they apply to. Consider now the set $\widetilde{\Delta}(\alpha\&\beta) = \widetilde{\Delta}(\alpha) \cup \widetilde{\Delta}(\beta)$ with the salience order that extends those of $\widetilde{\Delta}(\alpha)$ and $\widetilde{\Delta}(\beta)$ and is empty elsewhere. The membership order induced by the concept $\alpha\&\beta$ with associated defining feature set $\widetilde{\Delta}(\alpha\&\beta)$ is then exactly the order $\preceq_{\alpha\&\beta}$.

5.2 Categorial membership and concept determination

The *determination operator* can be used to account for the *modification* of a principal concept α by a modifier β. This determination, denoted by $\beta \star \alpha$, is most often translated in English by the combination of an adjective or an adjectived verb with a noun, like in the compositions *to-be-a-carnivorous-animal, to-be-a-flying-bird, to-be-a-french-student, to-be-a-red-apple*. However, it can also take the form of a noun-noun combination like in *to-be-a-pet-fish, to-be-a-barnyard-bird*, and, more generally, of a relative clause that will be globally encapsulated by the concept β (e.g.*to-be-an-American-who-lives-in-Paris*).

Note that the modifier becomes a *feature* of the compound concept $\beta \star \alpha$: *to-be-red* is clearly a feature of the composed concept *to-be-a-red-car*, and *to-be-a-woman* becomes a feature of the concept *to-be-physician-that-is-a-woman*.

It is important to keep in mind that we consider only the conceptual combinations that are *intersective*: the objects that fall under the composed concept $\beta \star \alpha$ are exactly the ones that both fall under α and under β. Thus, and to mention the most known examples, the determination connective cannot be used to form complex concepts like *to-be-a-brick-factory*, *to-be-a-criminal-lawyer* or *to-be-a-topless-district*: indeed, a brick factory need not be a factory that is made out of bricks, a criminal lawyer is not a lawyer that is a criminal, and a topless district is not a district that is topless (see (Kamp and Partee, 1995) for the distinction between intersective and non-intersective modifiers). Similarly, we shall not consider determination through features that quantify concepts, like in *to-be-a-tall-man* or *to-be-a-big-fish*, since these compositions are not intersective: a tall man is not something that is tall and that is a man, a big fish is not something that is big and that is a fish.

Observe finally that the star-operator $\star$ is a *partial* operator: given arbitrary α and β, it might be meaningless to try to form the concept $\beta \star \alpha$. For instance, there is no sense in talking of a *quick-number* or a *wooden-salience*. These pseudo-concepts correspond to nothing, and no object, real or fictitious, can be thought of falling under them, contrary to imaginary concepts like a *pink-elephant*, a *striped-apple* or a *flying-cow*: these latter definitely have a non-empty extension, because we can imagine a pink elefant, a striped apple or a flying cow. When forming the composition $\beta \star \alpha$, it is always understood that the intersection $Ext\,\alpha \cap Ext\,\beta$ is a non empty set.

Note also that unlike conjunction, the connective $\star$ is not supposed to satisfy commutativity: the concept of a *woman-that-is-a physcian* is not the same as the concept of a *physician-that-is-a-woman*(for examples and discussion, and in particular the distinction between *games-that-are-sports* and *sports-that-are-games* (see (Storms and al., 1999) with an ambiguous title however, mixing categorial membership and typicality) .

5.2.1 Evaluating membership relative to a modified concept

We suppose that α is a featured concept and that β is either a featured concept, or a simple feature whose associated membership order has been defined in the context of α. To extend the order $\preceq_\alpha$ (resp.$\preceq^*_\alpha$) to the concept $\beta \star \alpha$, giving preeminence to α over β, we set:

$$x \preceq_{\beta\star\alpha} y \text{ if } x \preceq_\alpha y \text{ and either } x \prec_\alpha y, \text{ or } x \preceq_\beta y.$$

Similarly, we define

$$x \preceq^*_{\beta\star\alpha} y \text{ if } x \preceq^*_\alpha y \text{ and either } x \prec^*_\alpha y \text{ or } x \preceq^*_\beta y.$$

In this model, the concept *to-be-a-flying-bird* will be considered as applying more to a penguin than to a bat, because the principal concept is that of *being-a-bird*: *to-fly* appears as a simple feature, less important than the concept it modifies. The relations we obtain this way are reflexive and transitive. They correspond to the membership orders induced by a featured concept whose defining feature set $\Delta(\beta \star \alpha)$ would include all the defining features of α and β, ordered with a salience order that extends the salience orders of $\Delta(\alpha)$ and of $\Delta(\beta) \setminus \Delta(\alpha)$, making furthermore any element of $\Delta(\alpha)$ more salient than any element of $\Delta(\beta) \setminus \Delta(\alpha)$.

Note that $\preceq_{\beta\star\alpha}$ and $\preceq^*_{\beta\star\alpha}$ are finite orders, so that it is possible to build from them a membership distance and a membership degree function, as was done in 2.1 and 4.3. Note also that this construction treats α and β as *features* of the new concept $\beta \star \alpha$. In fact, one easily checks that these compound membership orders exactly correspond to the lexicographic and the proportional preorders that would be induced by a fictitious featured concept whose defining feature set would simply consist in two 'features' α and β, α being considered as more salient than β.

In the particular case where β is itself a defining feature of α, the membership orders induced by $\preceq_{\beta\star\alpha}$ and $\preceq_\alpha$ agree: one has $x \preceq_{\beta\star\alpha} y$ iff $x \preceq_{\alpha y}$.

Membership relative to the concept *to-be-a-bird-with-wings* is the same than membership relative to the concept *to-be-a-bird*.

From its definition, the relation $\preceq_{\beta\star\alpha}$ is clearly a subrelation of $\preceq_\alpha$, and the same is true for $\preceq^*_{\beta\star\alpha}$ and $\preceq^*_\alpha$. This means that if an object y falls at least as much as x under the concept $\beta\star\alpha$, then it also falls at least as much as x under α.

Concerning the corresponding strict partial orders, it holds:

$$x \prec_{\beta\star\alpha} y \text{ if and only if either } x \prec_\alpha y, \text{ or } x \preceq_{\beta\star\alpha} y \text{ and } x \prec_\beta y$$

$$x \prec^*_{\beta\star\alpha} y \text{ if and only if either } x \prec^*_\alpha y, \text{ or } x \preceq^*_{\beta\star\alpha} y \text{ and } x \prec^*_\beta y.$$

As in the case of simple concepts, full categorial membership can be recovered through the $\preceq_{\beta\star\alpha}$-maximal or $\prec^*_{\beta\star\alpha}$-maximal elements. More precisely, we have the following result:

Observation 5 *Define the* extension $Ext\,(\beta\star\alpha)$ *of* $\beta\star\alpha$ *to be the set of all* $\preceq_{\beta\star\alpha}$*-maximal elements, and let* z *be an arbitrary object. Then the following conditions are equivalent:*

- z *is* $\prec_{\beta\star\alpha}$*-maximal*
- z *is* $\prec^*_{\beta\star\alpha}$*-maximal*
- $z \in Ext\,(\beta\star\alpha)$
- $z \in Ext\,\alpha \cap Ext\,\beta$
- $\forall x \notin Ext\,(\beta\star\alpha)$, $x \prec_{\beta\star\alpha} z$
- $\forall x \notin Ext\,(\beta\star\alpha)$, $x \prec^*_{\beta\star\alpha} z$.

Membership relative to modified concepts again appears to be *compostional*: the equality $Ext\,(\beta\star\alpha) = Ext\,\alpha \cap Ext\,\beta$ shows that an object fully falls under a modified concept if and only if it fully falls both under the concept and under its modifier. The category of *red-cars* exactly covers all the items that are red and that are cars.

5.3 Membership degree for compound concepts: the conjunction effect

The *conjunction effect* or *guppy effect* was observed in 1981 by Osherson and Smith (Osherson and Smith, 1981); it has been thereafter at the center of numerous research and experiments (see in particular (Hampton, 1988), (Kamp and Partee, 1995) or more recently (Aerts, 2009) and (Franco, 2009); it can be described by the fact that an item may be found to be more strongly a member of the composition of two concepts than a member of one of them. Thus, a cuckoo was found to be more strongly a member of the composed concept *(to-be-a-pet-bird)* than a member of the concept *to-be-a-pet* on its own, and a guppy more a member of the concept *(to-be-a-pet-fish)* than a member of the concept *to-be-a-fish*. This appears to be paradoxical, since any item falling under a composition $\alpha \star \beta$ must already fall under each of its components.

A similar effect, the *Linda paradox*, was observed in (Tversky and Kahneman, 1983): subjects were told about a woman, Linda, who has been involved at college in liberal politics. Some subjects were then asked to rate the probability that Linda has become a *bank teller,* other subjects were asked to rate the probability that she became a feminist bank teller. The result showed that it was judged more probable that Linda has become *a feminist bank teller*. This again seems in contradiction with classical logic and probability theory, since being a feminist bank teller necessarily implies being a bank teller, so that the probability of the first event should not exceed that of the second event. However, and in spite of '*desperate manipulations designed to induce subjects to obey the conjunction rule*', the result of the experiments all concluded in the sense of a so-called 'violation' of the conjunction rule.

These two 'paradoxes' seem to have promoted the introduction in concept theories of the quantum mechanics formalism: thus, Franco writes that '*Quantum mechanics, for its counterintuitive predictions, seems to provide a good formalism to describe puzzling effects of contextuality*'. Similarly, Aerts pleads for adopting in cognition theory the attitude of theoretical physicists for whom '*data showing deviations from set theoretic rules are a major indication of the presence of quantum structure*'; he thus devotes special attention to the example of the guppy effect as '*none of the currently existing concept theories provides a satisfactory descriptionand/or explanation of such effect for concept combinations.*'

We do not take sides on the question whether the rather complex and

artificial formalism of quantum mechanics is or is not suitable to model cognitive theories; however, we depart from the quoted authors in that we think that the guppy effect and the conjunction fallacy can be simply described and explained through a classical formalism, using for instance the tools which were developed in the preceding sections.

Since in the conjunction effect we compare the categorial membership of a same object relatively to two different concepts, we need to precisely evaluate this categorial membership. The only natural tools we can use for this purpose are those of *membership distance* and *membership degree*.

Let us denote by $\mu_{\beta\star\alpha}$ (resp. $\mu^*_{\beta\star\alpha}$) the membership distances corresponding to $\prec_{\beta\star\alpha}$ and $\prec^*_{\beta\star\alpha}$. Since $\prec_\alpha$ is embedded in $\prec_{\beta\star\alpha}$, a $\prec_\alpha$-chain is also a $\prec_{\beta\star\alpha}$-chain, and one has $\mu_\alpha(x) \leq \mu_{\beta\star\alpha}(x)$. Similarly, we have $\mu^*_\alpha(x) \leq \mu^*_{\beta\star\alpha}(x)$.

Recall that the associated *membership degrees* $\delta_{\beta\star\alpha}$ and $\delta^*_{\beta\star\alpha}$ are defined by

$$\delta_{\beta\star\alpha}(x) = 1 - \tfrac{\mu_{\beta\star\alpha}(x)}{N_{\beta\star\alpha}} \text{ and } \delta^*_{\beta\star\alpha}(x) = 1 - \tfrac{\mu^*_{\beta\star\alpha}(x)}{N^*_{\beta\star\alpha}},$$

where $N_{\beta\star\alpha}$ (resp. $N^*_{\beta\star\alpha}$) is the length of a maximal $\prec_{(\beta\star\alpha)}$ (resp.$\prec^*_{(\beta\star\alpha)}$)-chain. Note that $N_{\beta\star\alpha} \leq N_\alpha$, and $N^*_{\beta\star\alpha} \leq N^*_\alpha$.

Coming back to the guppy effect, we observe that, while the membership distance relative to a composed concept cannot be smaller than its distance relative to any of the components, the same is not true for the corresponding degrees: because of the presence of the normalizing constant $N_{\beta\star\alpha}$, we can have $\delta_\alpha(x) < \delta_{\beta\star\alpha}(x)$ for some item x, which will be therefore considered as falling more under $\beta\star\alpha$ than under α. This is exactly the guppy effect.

Example 6 *Let α be the concept* to-be-a-bird *and β the feature* to-be-black*, which for simplicity we suppose modelled by a simple two-valued function. The composition $\beta\star\alpha$ translates the concept* to-be-a-black-bird*. From Example 5, we know that the* birdhood degree *an agent may attribute to a macaque k is $\delta_\alpha(k) = 1/4$. Starting from the maximal α-chain*

$$m \prec_\alpha k \prec_\alpha b \prec_\alpha d \prec_\alpha s$$

of Example 5, the agent may consider the maximal $(\beta\star\alpha)$-chain

$$m \prec_{\beta\star\alpha} m' \prec_{\beta\star\alpha} k' \prec_{\beta\star\alpha} k \prec_{\beta\star\alpha} b \prec_{\beta\star\alpha} b' \prec_{\beta\star\alpha} d \prec_{\beta\star\alpha} d' \prec_{\beta\star\alpha} s \prec_{\beta\star\alpha} r,$$

where the items are

m*: a* white mouse*;* m'*: a* black cat*;* k'*: a* (red-brown) capuchin*;* k*: a* (black) macaque*;* b*: a* brown bat*;* b'*: a* black bat*;* d*: a* (blue) dragonfly*;* d'*: a* (black) fly*;* s*: a* sparrow*;* r *a* raven.

For this agent, it now holds $N_{\beta\star\alpha} = 9$ *and* $\mu_{\beta\star\alpha}(k) = 6$*, which leads to* $\delta_{\beta\star\alpha}(k) = 1/3$*: the* macaque *falls more strongly under the compound concept* to-be-a-black-bird *than under the component* to-be-a-bird *taken alone.*

Note that for an agent that would omit the link corresponding to k' *- for example an agent knowing only one species of monkeys - the length of a maximal chain would reduce to* $N_{\beta\star\alpha} = 8$ *and the resulting degree would then be* $\delta_{\beta\star\alpha}(k) = 1/4$*. In this case, no guppy effect would be observed. This agrees with the fact that the guppy effect is not universal, as shown in the experiments: some agents still consider that the guppy is no more a member of the concept* (to-be-a-pet-fish) *than of the concept* to-be-a-fish.

It can be objected that in the case of *full membership* our model seems to be contradicted by Hampton's conclusions (Hampton, 1988), following which *a large number of items ...(are) more often judged to belong in a conjunction such as* school furniture *or* protective clothing *than in the categories from which these concepts are supposedly drawn, namely* furniture *and* clothing. *Thus,* he pursues*, items belonged in a conjunction that did not belong in one of its constituent concepts.* This stands in total contradiction with the results of our Observation 5, by which which an object belongs to the extension of the compound $\beta\star\alpha$ if and only if it belongs both to the extension of α and to the extension of β. But if we look closely at the different experimental results displayed in Hampton's paper, we note that they are not really explicit as long as *full membership* is concerned. Let us recall indeed that membership, in Hampton experiments, was rated by a positive number from 1 to 3 'to indicate degree of typicality', while non-membership was rated by a negative number from -1 to -3 'to indicate relatedness as a non member'. In other words, Hampton treated on the same level categorial membership and typicality. It turns out that in the case of proper concept determination (thus excluding the cases of *kitchen furniture, sport vehicle* and *protective clothing*), the results, displayed pp.22-23 of Hampton's paper show that when an item's membership relative to one of the components is less or equal to -1, its membership relative to the conjunction is always strictly less than 1. This means that an item that is known not to fall under one of the components of a conjunction will never be considered as (fully) falling under this conjunction.

6 Structured defining sets

In developing our theory of elementary featured concepts, we restricted ourselves to the simple and rather isolated cases where the defining feature set consists of a simple list of elementary features. Usually however, definitions are of a more complicated type, as they rest on a set of key words - nouns, verbs, and modifiers - to which an apparatus consisting of auxiliary verbs, pronouns, locutions and ingredient markers provides the final *Gestalt*. Thus, the *grammar* of a definition is as important as the set of defining features it proposes. This fact has already been underlined by Ray Jackendoff (Jackendoff, 1993) and Anna Wierzbicka (Wierzbicka, 1996). In particular, in her research on a *Natural Semantic Metalanguage* (NSM), Wierzbicka and her followers proposed, together with a list of conceptual primitives, a list of conceptual elementary structures that constitute the syntax of this (meta)language ((For an introduction to Wierzbicka's work, see (Peeters and Goddard, 2006) or (Koselak, 2003)).

Categorial membership becomes more difficult to evaluate in the case of a concept that uses a non-trivial grammar for its definition, but the method developed in the preceding sections still remains valid for this elaborated case. To illustrate this, let us consider, as an example of resolution, the following (structured) definition of a *maple*:

'*Tall tree growing in northern countries, whose leaves have five points, and the resin of which is used to produce a syrup'.*

We first note that, among the features listed in this definition, only that of *tree* directly applies to the target concept *maple*. All the other ones are linked with some secondary concepts: thus, *tall* refers to *tree*, *to-have-fivepoints* refers to *leaves*, *northern* refers to *country*, and *syrup* refers to *resin*. The key features that intervene in this definition are the concepts *tree, tall, northern, fivepoints, syrup*. The apparatus is encapsulated by the sequence: **is a** + *key-feature*, **that is** + *key-feature*, **whose growing country is** +*key-feature*, **whose shape of leaves is** + *key-feature*, **whose resin produces** + *key-feature*. We can formalize the whole definition by a tree in which the edges translate the verbs used in the definition process, and the nodes underlined in roman letters stand for the key features while those in italics indicate the auxiliary concepts used in the *Gestalt*(see fig.1).

Suppose now we want to evaluate the categorial membership of a given object x relatively to the concept *to-be-a-maple*. This maple membership of

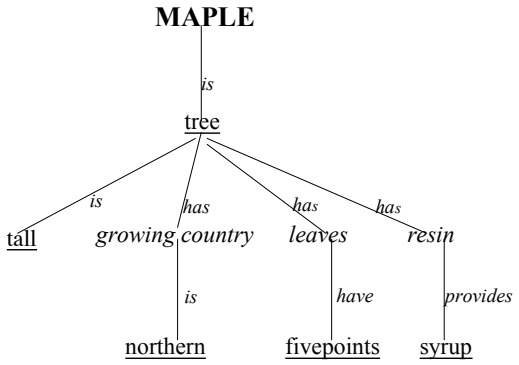


Figure 1: definition tree for *maple*.

x not only depends on its *tree* membership but also on the membership of *other object*s (eg. the *leaves* of x, the *resin* of x) relative to *auxiliary concepts* (eg.*to-have-five-points, to-provide-a-syrup*): it is not the object x itself that may be qualified as *having fivepoints* but the auxiliary object 'leaves of x'. We see in this example that there exists an important difference with the simple defining features sets used in the preceding sections, where the membership of an object relative to the target concept was directly evaluated through the membership of *this same object* relative to the defining features.

The problem is that the auxiliary objects that correspond to the auxiliary concepts may simply not exist: thus, it may be the case that x is not a tree, so that its *tallness* as a tree is meaningless; it may also happen that x is a pine-tree and has no leaves, in which case it is again meaningless to pretend to evaluate the membership of *its* leaves relative to the concept *to-have-fivepoints*.

To solve this problem, we have to re-interprete the initial definition by saying that an object is a maple if it is a tree *that* is tall, if this tree has a growing country *that* is northern and if this tree has resin *that* produces syrup. This translation enables us to consider the membership of a same single object

relative to the four compound concepts (*to-be-tall*)$\star$(*to-be-a-tree*), *to-have-a-*((*northern*)$\star$(*growing-country*)), *to-have*((*five points*) $\star$(*leaves*)), *to-have-a-*((*producing syrup*)$\star$(*resin*)).

The first concept - *(to-be-tall)$\star$(to-be-a-tree)* - is obtained through a determination, and its application to any object will be evaluated as shown in Section 5. As for the other ones, we have to consider them as complex *indecomposable* concepts: there exists indeed no way to compute their associated membership through that of their constituents, the same way we could not decompose the concept *to-eat-a-red-apple* using the concepts *to-eat*, *to-be-an-apple* and *to-be-red*. We retrieve here the difference of treatment between one-place and two-place predicates.

In order to compare the respective *maplehood* of two items, it is therefore enough to compare their respective categorial membership relative to each of the four mentioned concepts. This amounts to associating with the target concept *to-be-a-maple* the following set considered as a set of defining features

$\{$(*to-be-tall*)$\star$(*to-be-a-tree*), *to-have-a-*((*northern*)$\star$(*growing-country*)), *to-have*((*five points*) $\star$(*leaves*)), *to-have-a-*((*producing syrup*)$\star$(*resin*))$\}$.

It is again possible to equip this set with a salience order, yielding an associated membership order as was done in the case of featured concepts.

7 Conclusion

For a specific class of concepts, the *featured concepts*, we have shown that membership to the corresponding category can be easily modelled through a simple and natural preorder that takes into account the salience relation which may exist among the defining features of a concept, even in the case where this salience is not evaluated through a degree function. The proposed model easily extends to compound concepts, and explains the phenomenon of *conjunction effect* without appealing to the more elaborated and much sophisticated latest quantum theories. This model can also be used in the case of concepts whose 'definitions' are more sophisticated, adding to a list of features or key-words an elementary grammar that has to be taken into account.

ACKNOWLEDGMENTS

I wish to thank Daniel Lehmann for numerous advices and corrections that helped improving this work. I am also indebted to the anonymous referees for their numerous and helpful remarks and suggestions.

A Proofs

Observation 1 *Let $\preceq_\alpha$ be the relation defined by*

$x \preceq_\alpha y$ iff for each feature $f \in \Delta(\alpha)$ such that $y \prec_f x$, there exists a feature $g \in \Delta(\alpha)$, g more salient than f, such that $x \prec_g y$.

Then $\preceq_\alpha$ is transitive.

Proof: Let x, y and z be three items such that $x \preceq_\alpha y$ and $y \preceq_\alpha z$, and suppose that there exists $f \in \Delta(\alpha)$ such that such that $z \prec_f x$. We have to show that there exists a feature $g \in \Delta(\alpha)$, g more salient than f, such that $x \prec_g z$. We make a proof by cases:

- Suppose first that $x \preceq_f y$. Then we have $z \prec_f y$, and there exists a feature $k \in \Delta(\alpha)$, k more salient than f, such that $y \prec_k z$. We can suppose that k is maximally salient in $\Delta(\alpha)$ for this property (recall that $\Delta(\alpha)$ is finite). If $x \preceq_k y$, we get $x \prec_k z$ and we are done. Otherwise, because of the connectedness of $\preceq_k$, we have $y \prec_k x$. Since we supposed $x \preceq_\alpha y$, this implies that there exists a concept $g \in \Delta(\alpha)$, g more salient than k, such that $x \prec_g y$. We cannot have $z \prec_g y$ otherwise there would exist $h \in \Delta(\alpha)$, h more salient than g, such that $y \prec_h z$, contradicting the choice of k. We have therefore $y \preceq_g z$, hence $x \prec_g z$ as desired.

- Suppose now that $y \prec_f x$. Then there exists $k \in \Delta(\alpha)$, k more salient than f such that $x \prec_k y$, and we can again suppose k maximally salient for this property. If $y \preceq_k z$, we get $x \prec_k z$ and we are through. Otherwise, we have $z \prec_k y$ and there exists g more salient than k such that $y \prec_g z$. Let us show that $x \preceq_g y$: if this were not the case, we would have $y \prec_g x$, so that there would exist h more salient than g such that $x \prec_h y$. But then h would be more salient than k, which is impossible. We have therefore $x \preceq_g y$, hence $x \prec_g z$, and the proof is complete. ■

Observation 2 *One has $x \sim_\alpha y$ if and only if $x \sim_f y$ for all features f of $\Delta(\alpha)$.*

Proof: It is clear that $x \sim_\alpha y$ holds whenever $x \sim_f y$ for all features f of $\Delta(\alpha)$. Suppose conversely that we do not have $x \sim_f y$ for some feature f of $\Delta(\alpha)$, and let f be of maximal salience in $\Delta(\alpha)$ for this property. Since $\preceq_f$ is a total preorder, we necessarily have $x \prec_f y$ or $y \prec_f x$. In the first case, and by the choice of f, we cannot have $y \preceq_\alpha x$; in the second case, we cannot have $x \preceq_\alpha y$. This shows that we cannot have $x \sim_\alpha y$. ■

Observation 3 *Define the* extension *$Ext\,\alpha$ of α as the set of all objects z such that $\delta_f(z) = 1\,\forall f \in \Delta(\alpha)$. Then it holds $x \prec_\alpha z$ and $x \prec^*_\alpha z$ for all $x \notin Ext\,\alpha$ and $z \in Ext\,\alpha$.*

Proof: If z is such that $\delta_f(z) = 1$ for all $f \in \Delta(\alpha)$, it follows from the definitions of $\preceq_\alpha$ and $\preceq^*_\alpha$ that one has necessarily $x \preceq_\alpha z$ and $x \preceq^*_\alpha z$ for all objects x of the universe. If moreover there exists a defining feature f such that $\delta_f(x) < 1$, these inequalities are strict, so that $x \prec_\alpha z$ and $x \prec^*_\alpha z$ as desired. ■

Observation 4 *. $\preceq_\alpha$ and $\preceq^*_\alpha$ are finite membership orders.*

Proof: Concerning $\preceq_\alpha$, the proof follows from Observation 2. As for $\preceq^*_\alpha$, we see, again by Observation 2 that every$\preceq_\alpha$- equivalence class is embedded in a $\preceq^*_\alpha$-equivalence class, whence the result.

Observation 5 *Define the* extension *$Ext\,(\beta \star \alpha)$ of $\beta \star \alpha$ to be the set of all $\preceq_{\beta\star\alpha}$-maximal elements, and let z be an arbitrary object. Then the following conditions are equivalent:*

- $z \in Ext\,(\beta \star \alpha)$
- *z is $\prec^*_{\beta\star\alpha}$-maximal*
- $z \in Ext\,\alpha \cap Ext\,\beta$
- $x \prec_{\beta\star\alpha} z\ \forall x \notin Ext\,(\beta \star \alpha)$
- $x \prec^*_{\beta\star\alpha} z\ \forall x \notin Ext\,(\beta \star \alpha)$.

Proof: Since we have $\prec^*_{\beta\star\alpha} \subseteq \prec_{\beta\star\alpha}$, any $\prec_{\beta\star\alpha}$-maximal element is $\prec^*_{\beta\star\alpha}$-maximal. If x is not $\prec^*_{\beta\star\alpha}$-maximal, it holds $x \prec^*_{\beta\star\alpha} z$ for any element z of $Ext\,\alpha \cap Ext\,\beta$, since $\prec^*_{\beta\star\alpha}$ corresponds to the proportional order induced by the fictitious defining features set $\{\alpha, \beta\}$. This shows that any $\prec^*_{\beta\star\alpha}$-maximal element must lie in $Ext\,\alpha \cap Ext\,\beta$, and is therefore also $\prec_{\beta\star\alpha}$-maximal. The proof of the Observation follows. ■

References

Aerts, D. (2009). Quantum structure in cognition. *Journal of Mathematical Psychology*, (53-5):314–348.

Arrow, K. J. (1953). *Social Choice and Individual Values*. Cowles Foundation Monographs; New York: Wiley 1964.

Dubois, D., Esteva, F., Godo, L., and Prade, H. (2005). An information-based discussion of vagueness. In Cohen and Lefebvre, editors, *Handbook of Categorization in Cognitive Science*, pages 892–913.

Fodor, J. (1998). *Concepts: where cognitive science went wrong*. Oxford University Press.

Franco, R. (2009). The conjunction fallacy and interference effects. *Journal of Mathematical Psychology*, (53-5):415–422.

Frege, G. (1879). *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle.

Freund, M. (2008). On the notion of concept I. *Artificial Intelligence*, (172):570–590.

Freund, M. (2009). On the notion of concept II. *Artificial Intelligence*, (173):167–179.

Gärdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought*. MIT Press.

Hampton, J. (1988). Overextension of conjunctive effects: Evidence for a unitary model of concept typicality and class inclusion. *Journal for Experimental Psychology : Learning, Memory, and Cognition*, (14):12–32.

Hampton, J. (1993). Prototype models of concept representation. In Van Mechelen et al.(Eds.), L. A. P., editor, *Categories and concepts: Theoretical views and inductive data analysis*.

Hampton, J. (1995). Testing the prototype theory of concepts. *Journal of Memory and Laguage*, (34):686–708.

Jackendoff, R. (1993). *Patterns in the mind: language and human nature*. New York Harvester Wheatsheaf.

Kamp, H. and Partee, B. (1995). Prototype theory and compositionality. *Cognition*, (57):129–191.

Koselak, A. (2003). La semantique naturelle d'Anna Wierzbicka et les enjeux interculturels. *Questions de communication*, (4):83–95.

Lee, J. (2003). Ordinal decomposability and fuzzy connectives. *Fuzzy sets and systems*, (136):237–249.

Lehmann, D. (2010). personal communication.

Nardi, D. and Brachman, R. (2003). An introduction to description logics. In Baader, F., editor, *The Description Logic Handbook*, pages 1–44. Cambridge University Press.

Osherson, D. and Smith, E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, (11):237–262.

Osherson, D. and Smith, E. (1982). Gradedness and conceptual combination. *Cognition*, (12)):299–318.

Peeters, B. and Goddard, C. (2006). The natural semantic metalanguage (nsm) approach: An overview with reference to the most important Romance languages. *In B. Peeters (ed.), Semantic primes and universal grammar. Empirical evidence from the Romance languages*, (19):13–88.

Putnam, H. (1975). The meaning of meaning. In *Mind, language and reality*, pages 115–120. Cambridge University Press.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology*, (104):192–233.

Rosch, E. and Mervis, C. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, (7):573–605.

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

Smith, E., Shoben, E., and Rips, L. (1974). Structure and process in semantic memory: a featural model for semantic decisions. *Psychological Review*, (81):214–241.

Tversky, A. and Kahneman, D. (1983). Extension versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review*, (90):141–168.

Wierzbicka, A. (1996). *Semantics. Primes and Universals*. Oxford University Press.

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, (8):338–353.

Zadeh, L. (1982). A note on prototype theory and fuzzy sets. *Cognition*, (12):291–297.