Proceedings of Sinn und Bedeutung 11, E. Puig-Waldmüller (ed.), Barcelona: Universitat Pompeu Fabra, pp.509-524.

THE LOGIC OF THE TYPICALITY JUDGMENTS

Galit W. Sassoon,
Tel Aviv University

gala@post.tau.ac.il

Abstract

The standard model theoretical view of word meaning is incompatible with known facts about natural categories ('the typicality effects'), which experimental studies in cognitive psychology robustly support. Parts 1-3 of this paper present and criticize Kamp and Partee's 1995 well known analysis of the typicality effects. The main virtue of this analysis is in the use of supermodels, rather than fuzzy models, in order to represent vagueness in predicate meaning. The main problem is that typicality of an item in a predicate is represented by a value assigned by a measure function, indicating the proportion of supervaluations in which the item falls under the predicate. A number of issues cannot be correctly represented by the measure function, including the typicality effects in sharp predicates; the conjunction fallacy; the context dependency of the typicality effects, etc. In Parts 4-5, it is argued that these classical problems are solved if the typicality ordering is taken to be the order in which entities are learnt to be denotation members (or non-members) through contexts and their extensions. A modified formal model is presented, which clarifies the connections between the typicality effects, predicate meaning and its acquisition.

1 What are the typicality effects?

The structure of the natural categories which form the denotations of nouns like bird was extensively investigated by cognitive psychologists over the past forty years. The empirical findings unequivocally show that the natural categories possess a gradable structure. These findings are called 'the typicality effects'. The most basic typicality effects are the findings that speakers order entities or subkinds by typicality, i.e., by how good an example, representative, (proto)typical or related to the category each one of them is. For example, a robin is often considered more typical of a bird than an ostrich or a penguin. In addition, a bat or a butterfly is often considered more related or more similar to a bird than a dog or a cow.

Normally, these effects are understood as psychological in nature and hence not the direct concern of the linguist. We will see, however, that a better understanding of the typicality effects is required in order to improve our understanding of central linguistic issues. For example, one question of interest to the linguist is whether the noun meaning is gradable or not. Predicative comparative relations are felicitous only if the meaning of their predicative argument is gradable. For example, sentence (1) is felicitous because the predicative argument *tall* denotes a gradable category, but sentence (2a) is odd, and this fact is generally accounted for by postulating that the meaning of *bird* is not gradable (Kennedy & McNally 2005).

(1) *Robins are* ***taller / less tall*** *than ostriches*

(2)

a. # *Robins are* ***more / less birds*** *than ostriches*

b. # *Number 31 is more / less prime than number 3*

The problem is that a graded structure can be indirectly accessed in many different ways. In (3a) we see that entities can be indirectly ordered by how *good an example* they are of *bird*.

Given (3a) and (2a) alone, we could reasonably conclude that a graded structure is indeed related to the concept BIRD, but not necessarily to the meaning of the lexical entry *bird*. In (3b) we see that, when modified by modifiers like *'typical'*, the noun *bird* becomes felicitous in the comparative. How can that be the case, if the meaning of *bird* is inherently non-gradable? We can speculate that the adjective *typical* functions like the adjective *likely* or *probable*. If the adjective *typical* links each noun with some type of probability scale, then example (3b) is no mystery at all. It is, indeed, frequently suggested in the literature that the typicality ordering judgments stem from objective membership probabilities.

(3)

a. *Robins are* ***better examples of birds*** *than ostriches*
b. *Robins are* ***more typical birds*** *than ostriches*

However, this idea is challenged by a variety of empirical findings. Obviously, for no item can the objective probability of membership in a conjunctive category (like *brown apple*) be greater than the objective probability of membership in each conjunct (*brown* or *apple*), which seems to make the prediction stated in (4a). But that means that theories of typicality based on objective probability (such as the Fuzzy models of Zadeh 1965; Lakoff 1973; Osherson & Smith 1981) make the same prediction for the typicality degrees, as stated in (4b). However, these predictions are not borne out by the data. First of all, speakers' judgments of membership probability (or likelihood) do not correspond to objective probabilities. For example, as stated in (5), speakers often believe that SOME items (like brown apples) ARE *more likely brown apples than apples*. This belief is called *the conjunction fallacy* (Tversky and Kahneman 1983). Second, thinking of typicality as membership-probability also renders fallacious the intuition in (6), namely the intuition that SOME items (the brown apples again) ARE *more typical* in *brown apple* than in *apple*. This is called *the conjunction effect* (Smith et al 1988). Third, even the basic judgment in (7c), namely the belief that SOME items (ostriches) ARE *more typical* in *ostrich than* in *bird*, becomes a fallacy for similar reasons, because, as stated in (7a), for no item can the likelihood of being *a bird of a certain type* exceed that of being *a bird of whatever type*. Let us call this *the subtype effect*.

(4)

a. Prediction: NO items are *more likely brown apples than apples*.
b. Prediction: NO items are *more typical in 'brown apples' than in 'apples'*.

(5) *The conjunction fallacy*: SOME items ARE more likely brown apples than apples.
(6) *The conjunction effect*: SOME items ARE *more typical* in *brown apple* than in *apple*.
(7)

a. Prediction: NO items are *more likely ostriches than birds*.
b. Prediction: NO items are *more typical in 'ostrich' than in 'bird'*.
c. *The subtype effect:* SOME items ARE *more typical* in *ostrich than* in *bird*.

Given these intuitions, speakers' probability judgments are often called *subjective*. Can typicality be equated with subjective probability? Naturally occurring examples and experimental results show that the answer is no. Even speakers' subjective probability judgments are often dissociated from their typicality judgments. For instance, example (8) naturally occurred in the World Wide Web. It demonstrates that certain things can be judged both frequent in and atypical of a category. Furthermore, Teigen and Keren's 2003 experimental results show that surprise judgments are sometimes linked to typicality judgments but not to judgments of probability or expectedness. Thus, typicality can be equated neither with objective nor with subjective probability. Finally, the example in (7) shows that no special adjectival modification is required in order to access the graded structure related to *bird*. The use of the particle *of* in (9a), or the adjectival suffix '*y*' in (9b-c), is sufficient to turn the meaning of the word *bird* gradable.

(8) *axat ha-tofaot ha-yoter-s'xixot ve-ha-lo-tipusiot hi kocer nes'ima bemaa'mac lelo muaka baxaze.* ('one of the more frequent and non-typical phenomena [related to angina pectoris] is shortness of breath during exertion without discomfort in the chest')

(9)

a. *A robin is* ***more of a bird*** *than an ostrich*

b. *The noun 'activity' is* ***"nounier" / less "nouny"*** *than the noun 'bird'*

c. *Number 31 is* ***less "primy"*** *than number 3*

Thus, we must conclude that the grammar links the so-called 'non-gradable' (or 'sharp') nouns with some meaning features which enable us to turn them gradable without difficulty, when appropriately modified. Nouns, like adjectives, are inherently associated with entity orderings (scales), which do not correspond to probability scales. A proper linguistic account of the meaning of nouns needs to say what these orderings are, and why they can only be denoted by a noun which is modified, as demonstrated in examples (2), (3) and (9). This account should also explicate the semantic and pragmatic differences and similarities between different comparative relations, such as *more, more of, more typical of, more relevant to, more likely, more similar,* etc.

Semanticists generally assume that the meaning of gradable adjectives like *tall* and *bald* includes an ordering dimension, or a feature, such as *height* in the case of *tall* (see, for instance, Kamp 1975; Kennedy & McNally 2005). As for nouns, experimental results established that speakers characterize them by *a rich set of features*. For example, the noun *bird* is characterized by features like *feathers, small, flies, sings, perches in trees, eats insects*, and so on (Rosch 1973). The notion *'feature'* is not well understood. The features which people link with predicates do not stand for necessary conditions for membership in the denotation (Wittgenstein 1968 [1953]). The more typical birds are generally more typical in the *bird* features, but, again, they often do not satisfy some of the features. Despite its being ill defined, the notion of a *feature* is crucial in accounts of unconscious processing effects which are connected to the basic typicality ordering judgments (cf. Smith, Shoben and Rips 1974). Typicality correlates with *online categorization time*. For example, verification time is shorter for sentences like *a robin is a bird* than for sentences like *an ostrich is a bird* (i.e., categorization time is shorter for typical instances; Rosch 1973). The proposal I present in this paper is compatible with, and in fact, complementary to, feature-based processing accounts, in that I propose a simple definitional constraint for feature-hood.

The typicality judgments are highly culture, language and context dependent. First, in an empirical investigation of second language typicality ratings, Malt and Sloman (2003) found that the typicality judgments (and also category membership judgments) of the least experienced learners diverged substantially from native responses, and even the most experienced learners retained some discrepancies from native patterns. Second, the knowledge about the typicality ordering is often partial even among the native speakers of a language. For example, if one bird sings and the other nests, which one is more typical? We cannot tell, out of context. The knowledge about the features can also be partial. For example, is the property *in the home* typical of *chairs*? Again, we cannot tell, out of context. Third, within a context of an utterance like (10), a chicken (not a robin) is regarded as a typical bird, and categorization time is faster for the contextually appropriate item *chicken*, not for the normally typical but contextually inappropriate item *robin* (Roth and Shoben 1983).

(10) *The bird walked across the barnyard*

Finally, often, within a context of utterance, speakers can compose *on the fly* meanings for complex expressions like *things to take from home in case of a fire* that were never stored in their memory before (Barsalou 1983). These composed meanings are often called '*ad-hoc concepts*', so as to emphasize this fact. Recently, focusing on basic lexical items, scholars argued that pragmatic effects, such as denotation narrowing or widening within a context of utterance, are also processes in which new ad-hoc concepts are created on the fly (Sperber and Wilson 1998). Taken together, these findings suggest that, within each context, speakers productively use a set of rules and strategies to build a typicality ordering relation and feature-set for both simple and complex concepts. These orderings and features are deeply connected to the denotations and to the ways they can be restricted or stretched.

The psychological findings which were reviewed so far have been replicated time and again. Yet, *the mental models underlying them* and *their relations to predicate meaning* are still a puzzle. To see this, we will now review the well known typicality theory, which is most frequently cited by formal semanticists, namely – *The Supermodel Theory*, by Kamp and Partee 1995. Its main innovation is in the use of a logic *with three truth values* and the technique of *Supervaluations* (van Fraassen 1969; Kamp 1975; Veltman 1984; Landman 1991), as opposed to the standard use of a logic with *multiple truth values* (such as *fuzzy logics*) in the analysis of typicality in artificial intelligence, cognitive psychology, and linguistics (for a review of such accounts and their problems see Zadeh 1965; Osherson and Smith 1981 and Kamp and Partee 1995). However, we will see in Part 3 that in some respects (which also characterize, to a large extent, earlier accounts) the supermodel analysis is limited and problematic. Among them is the use of membership functions, prototypes and distance functions. In Parts 4-5, a novel analysis is proposed which completely abandons these notions.

2 The supermodel theory (Kamp and Partee 1995)

A supermodel M^* consists of one partial model M, which I will call 'context' M. In M, denotations are only partially known. For example, the denotation of *chair* in a partial context M may consist of only one item – the prototypical chair, p_{chair}. The denotation of *non-chair* may consist of only one item too, which is very clearly not a chair, say – the prototypical sofa, p_{sofa}. This means that in M, we don't yet know if anything else (an armchair, a stool, a chair with less than 4 legs, without a back, of abnormal size, which is not used as a seat, etc.) is a *chair* or not. In addition, M is accompanied by a set T of total models (*supervaluations* in van Fraassen 1969), i.e., all the possibilities seen in M to specify the complete sets of *chairs* and *non-chairs*. In each t in T, each item is either in the denotation of *chair* or in the denotation of *non-chair*.

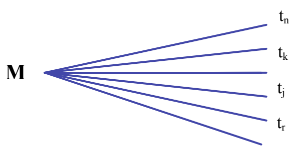

Figure 1: The context structure in a supermodel M^*

Formally, a supermodel M^* for a set of predicates A and a universe D is a tuple **<M,T,m>** s.t.:

(A) **M** is a *partial model*:

Predicates are associated with partial denotations in M: $\langle [P]^+_M, [P]^-_M \rangle$. For example, if $[chair]^+_M = \{d_1\}$ and $[chair]^-_M = \{d_3\}$, then d_2 is in the gap. We do not yet know if it is a chair or not. Non-vague predicates like *bird* are assumed to have no gap: $[bird]^+_M \cup [bird]^-_M = D$.

(B) **T** is *a set of total models* (completions of M):

In each t in T, the predicates are associated with disjoint and total denotations ($[P]^+_t \cap [P]^-_t = \varnothing$ and $[P]^+_t \cup [P]^-_t = D$), which are monotonic extensions of their denotations in M ($[P]^+_M \subseteq [P]^+_t$ and $[P]^-_M \subseteq [P]^-_t$). For example, in each t in T, d_2 is added to either $[chair]^+_t$ or $[chair]^-_t$.

Given this ontology, the *membership degree* of an individual d in a vague noun like *chair* is indicated by the size, or *measure*, of the set of total contexts in which d is a *chair*, $\mathbf{m}(\{t \in T: d \in [chair]^+_t\})$. For example, the prototypical chair, p_{chair}, is a *chair* in all total possibilities, so its membership degree is 1. The prototypical sofa, p_{sofa}, is a *chair* in no possibility, so its membership degree is 0. If a stool d is a *chair* in a third of the cases, its degree is 1/3, etc.:

(C) **m** is *a measure function* from sets of total models to real numbers between 0 and 1:

i.e., m is a function which satisfies the following constraints (Kamp and Partee 1995, p. 153): $m(T) = 1$; $m(\{\}) = 0$; $\forall T_1, T_2$, s.t. $T_1 \subset T_2$: $m(T_2) = m(T_1) + m(T_2 - T_1)$, etc.

(D) The *membership-degree* of d in P, $\mathbf{c_{m(d,P)}}$, is given by the measure m of the set of total models in which d is P: $c_{m(d,P)} = m(\{t \in T: d \in [P]^+_t\})$. E.g. $1 = c_{m(d1,chair)} > c_{m(d2,chair)} > c_{m(d3,chair)} = 0$.

(E) Predicates P are also associated with a *prototype* p (the best possible P) and a *typicality degree function*, which assign to each entity d a value, $c_{P(d,P)}$, between 0 and 1: d's distance from P's prototype. Predicates like *tall* or *odd number*, unlike *bird, grandmother, red,* etc., have no prototype (there is no maximal *tallness* or *oddness*).

How are the values of the typicality degree function calculated? Generally, they are given by the values of the membership function: $c_P \cong c_m$. For example, in *chair*, the more typical entities fall under $[chair]^+$ in more of the total models t in T. However, in predicates like *bird*, unlike predicates like *chair*, typicality (c_P) and membership (c_m) are separated, not coupled (or in Kamp and Partee's terminology, *the prototype does not affect the denotation*).

There are at least two reasons for this: First, intuitively, an *ostrich* d is a *bird* even in M, i.e., $c_{m(d,bird)} = 1$, but it is an atypical *bird*, i.e., $c_{P(d,bird)} < 1$. Thus, $c_m \neq c_P$. Second, intuitively, an *ostrich* is always a *bird*, i.e., for any entity d, the set of total contexts in which d is an *ostrich*, $\{t \in T: d \in [ostrich]^+_t\}$, is always a subset of the set of total contexts in which d is a *bird*, $\{t \in T: d \in [bird]^+_t\}$. So, $c_{m(d,ostrich)}$ is always smaller than $c_{m(d,bird)}$ as demonstrated in (11a). But recall *the subtype-effect*. Intuitively, d can be *more typical* of an *ostrich* than of a *bird*, so $c_{P(d,ostrich)}$ can be greater than $c_{P(d,bird)}$, as demonstrated in (11b). Again, $c_m \neq c_p$.

(11)
a. $c_{m(d,\,ostrich)} = m(\{t \in T: d \in [\,ostrich]^+_t\}) \quad \leq \quad m(\{t \in T: d \in [\,bird]^+_t\}) = c_{m(d,\,bird)}$
b. $c_{P(d,\,ostrich)} \quad > \quad c_{P(d,\,bird)}$.

We see that the subtype effect forces Kamp and Partee to separate membership and typicality.

Similarly, in any total context in which an entity d is a *brown apple,* d is an *apple*, i.e., the set

$\{t \in T: d \in [\textit{ brown apple}]^{+}_{t}\}$ is always a subset of the set $\{t \in T: d \in [\textit{apple}]^{+}_{t}\}$. Hence, the membership degree of d in *brown apple* can maximally reach d's degree in *apple* and not more as demonstrated in (12a). Thus, the membership function $c_{m(d,P)}$ cannot represent the *conjunction effect* and *fallacy*, e.g., the intuitive judgments that a brown apple is regarded as *more typical* (12b), as well as *more likely* (12c), *a brown apple* than *an apple*:

(12)

a. $c_{m(d,\textit{brown apple})} = m(\{t \in T: d \in [\textit{brown apple}]^{+}_{t}\}) \quad \leq \quad m(\{t \in T: d \in [\textit{apple}]^{+}_{t}\}) = c_{m(d,\textit{apple})}$

b. $c_{P(d,\textit{ brown apple})} \quad > \quad c_{P(d,\textit{ apple})}$

c. $\textit{likely}_{(d,\textit{ brown apple})} \quad > \quad \textit{likely}_{(d,\textit{ apple})}$

However, Kamp and Partee observe that the interpretation of modifiers like *brown* depends on the local context created by the noun they modify. For example, *brown* is interpreted differently when applied to *apple*, *skin, shelf, dress*, etc. Thus, Kamp and Partee propose to replace c_m in modified nouns like *brown apple* by a new function, which may assign d a higher value than $c_{m(d,\textit{apple})}$ or $c_{m(d,\textit{brown})}$. *The modified membership function* for the modified noun *brown apple*, $\mathbf{c_{m(d,\textit{brown /apple})}}$ is given by d's degree in *brown*, $m_{(d,\textit{brown})}$, minus 'a' – the minimal *brown* degree that the measure function m assigns to an apple. This value is normalized by the distance between 'a' - the minimal - and 'b' - the maximal - *brown* degrees assigned to apples. This normalization procedure ensures that the result ranges between 0 and 1:

(F) *The modified membership function for modified nouns:*
Let a and b be the minimal and maximal *brown* degrees among the apples in M, respectively:
$\mathbf{c_{m(d,\textit{brown /apple})}} = (m_{(d,\textit{brown})} - a) / (b - a)$.

For example, a brown apple d may score 0.9 in *brown*; the apple which is least brown may score 0 in *brown* (because some apples are not *brown* at all) and the apple which is most *brown* may score 0.95 (assuming that no apple is maximally brown). Then the value $\mathbf{c_{m(d,brown /apple)}}$ would be 0.974. This value indeed exceeds d's degree in *brown* (0.9) and possibly also d's degree in *apple*, as desired (13). Thus, conjunction falacies are represented using the modified membership function, and conjunction effects are represented, given that typicality in vague predicates is coupled with membership.

(13) $\mathbf{c_{m(d,brown /apple)}} = (0.9 - 0) / (0.95 - 0) = 0.974$.

3 Problems

3.1 Typicality degrees of denotation members

We will now see that the idea that measure-functions which range over total contexts can represent typicality has some fundamental problems. The first and most basic problem is that the measure function fails to account for the fact that denotation members are not necessarily associated with the maximal degree of typicality, 1. This is particularly problematic in non-vague nouns (the so-called 'sharp' nouns), which form the most prominent examples of the prototype theory. For example, even atypical birds like *ostriches* and *penguins* are known to be *birds*. That is, already in M, they are considered part of $[\textit{bird}]^{+}_{M}$. Indeed, the *bird* denotations are assumed to be completely specified, or in other words, not to vary across different total contexts. This is the standard way to represent the fact that predicates like *bird* are not – or are much less – vague than predicates like *chair* or *tall*. However, this is also the reason for which the measure function cannot indicate typicality in sharp predicates. Given that they are always known to be birds, the membership degree of atypical examples like *ostriches* and *penguins* in *bird* (i.e., the measure of the set of total contexts in which they are *birds*) is always 1. And for non-birds – be they *butterflies* and *bats* or *stools* and *cows* – since they are members in $[bird]^{-}_{M}$, their membership degree in *bird* is always 0. Intermediate typicality degrees in sharp nouns cannot be calculated using c_m, and Kamp and Partee do not specify how exactly typicality degrees (c_P values) are determined when c_m and c_P are separated.

3.2 The subtype and conjunction effects

Furthermore, we already saw that the measure function m fails to predict *the subtype effect*, e.g., the intuition that the typicality of ostriches in *ostrich* exceeds their typicality in *bird*. A membership degree (or measure m) is never bigger in *ostrich* than in *bird*, because in any total context in which an entity is an *ostrich*, it is also a *bird*. Nor can the modified membership function, which Kamp and Partee add to the model in order to capture the *conjunction fallacy* and *effect*, help us here. Why? Because the minimal and maximal *ostrich* degrees in $[bird]^{+}_{M}$ are 0 and 1. We can find both complete ostriches (of membership degree 1) and complete non-ostriches (of membership degree 0) among the birds. Consequently, $c_{m(d,ostrich\,/\,bird)}$ is identical to $c_{m(d,ostrich)}$ (14). Thus, we have to keep c_m and c_P separated in such lexical nouns. The values of c_P have to represent the subtype effect in *bird*. But then again, Kamp and Partee do not say how exactly c_P is fixed when it is separated from c_m. Hence, the subtype effect is not accounted for, and in addition to this, the separation between c_m and c_p forces us into an inelegant theory, which stipulates as primitives two unconnected sets of values for c_m and c_p.

(14) $c_{m(d,ostrich\,/\,bird)} = (m_{(d,ostrich)} - 0) / (1 - 0) = c_{m(d,ostrich)}$

Conjunction fallacies in modified nouns are also not dealt with correctly. Indeed, brown apples are allowed to have greater degrees in *brown apple* than in *brown* or in *apple*, as desired, but they are ordered only by how *brown* they are. This yields incorrect degrees. Intuitively, an apple of an unusual shape or size, which is therefore assigned, say, typicality degree 0.2 in *apple*, even if maximally *brown* (degree 1), is considered an atypical *brown apple* (Smith et al 1988), not a maximally typical *brown apple* (of degree 1), as predicted (15). Many naturally occurring examples like (16)-(18) which refer to typicality in negated and modified nouns can be found in a simple Google search. Hence, we cannot dismiss the problems in predicting typicality in complex predicates on the basis that typicality is inherently non-compositional or non-productive. We need an analysis which will correctly predict speakers' intuitions about typicality in complex predicates when such intuitions exist.

(15) $c_{m(d,brown\,/apple)} = (m_{(d,brown)} - a) / (b - a) = (1 - 0) / (1 - 0) = 1$
(16) *... pretty much typical of a non-fan, non-entertainment, smart up-market British paper*
(17) *What were some exercises you would do on a typical non-running day?*
(18) *You counter with an anecdotal tale about a non-typical non-developer. How does your counter-argument apply to a typical non-developer?*

3.3 Partial knowledge and numerical degrees

Another classical problem concerns the representation of context dependency in the typicality judgments. This problem has to do with the fact that the membership (measure) functions are total (m assigns to every entity a degree in every predicate in M), though knowledge about typicality is often partial. Which bird is more typical – an ostrich or a penguin? A bird that sings or a bird that nests? Many contexts are too partial to tell such facts. Nor do speakers know every typicality feature in every partial context. For example, is *in the home* typical of chairs? The representation of knowledge about typicality in M must be partial and context (valuation) dependent. But the measure function is defined per supermodel (for each entity, it is a measure of the proportion of valuations in T in which it is a predicate member), so it is not easy to see how this measure function can be relativized to a valuation.

A related problem is that numerical degrees are not intuitive primitives. For example, why would a certain penguin have a degree 0.25 rather than, say, 0.242 in *bird*? Kamp and Partee suggest solving this problem by associating M with a set of measure functions, such that we would only know in M that, e.g., the degree of a penguin is between 0.25 to 0.242 in *bird*. However, Kamp and Partee admit that this solution is complex and not quite natural. It is true that in the languages of the world the comparative form *more / less P than* is derived from the predicate form *P*, which, presumably, stands for the concept *P to degree* μ (Kamp 1975). Nevertheless, at least as far as typicality is concerned, representing the typicality ordering which is denoted by a typicality comparative (e.g., the intuition that penguins are *less typical than* ducks, which in turn are *less typical than* robins, etc.), and deriving the degrees from this ordering by some general strategy seems to be a more intuitive approach.

3.4 Prototypes versus feature sets

The notion of an entity prototype is problematic in that it is exceptionally unfruitful when it comes to compositionality, i.e., in predicting prototypes of complex concepts from the prototypes of their constituents (Kamp and Partee 1995; Hampton 1997). Consider negations: What would the prototype of *non-bird* be: a dog, a day, a number? Similarly for conjunctions: What would the *male-nurse* prototype be, given that a typical *male-nurse* may be both an *atypical male* and an *atypical nurse* (ibid). Another problem has to do with predicates which are lacking a prototype. For example, there is no maximum tallness. But with no prototypes, the intuition that there are typical (and atypical) *tall players, tall teenagers, tall women*, etc., is not accounted for. The status *prototypical*, so it seems, ought to be given to an entity only within a context (a valuation); there are no context-independent entity-prototypes.

Finally, the prototypes force us into a complicated taxonomy of predicate types, with different mechanisms in their meaning: With or without a prototype; with a prototype that affects the denotation or that does not affect the denotation; with a vague or a non-vague meaning, etc. This is especially problematic when compositionality is addressed (Kamp and Partee 1995). For example, of what type are conjunctions of different predicate types, like *tall bird*, where *tall* is a vague predicate without a prototype, and *bird* is a non-vague predicate with a prototype? Furthermore, the obligatory modification by *typical-of* in the comparative cannot mark the presence or absence of a prototype, because some predicates have a prototype but occur in comparatives without *typical of* (e.g., *red*), while others have no prototype but must (e.g., *non-bird*) or can (e.g. *tall thing*) occur with *typical of* in comparatives.

The main purpose of entity prototypes is to avoid the notion of *feature-sets*, which Kamp and Partee see as an ill-defined notion. Feature-based models are most widespread in the analysis of typicality (Rosch 1973; Smith et al 1988). The main idea in feature models is what I call *The Weighted Mean Hypothesis*. According to this hypothesis, the typicality degree of, e.g., the subkind *robin* in *bird* is indicated by the weighted mean of all the degrees of *robin* in the *bird* features: How well it scores in *flies*, *sings*, *small*, etc. Indeed, empirical studies show that speakers must have some capacity to average over features. However, there is little agreement about the ways entities' degrees in the features are determined, about the ways features are chosen and assigned weights (degree of importance), and about the precise averaging method. Thus, features alone do not form a sufficient account. But then again, avoiding feature-sets is also problematic. One can only categorize novel instances on the basis of their distance or similarity to a known prototype if there is some means of determining the connections that exist between the instances and the prototype (Hampton 1997). And for this reason too, theories without feature sets fail to predict the connections between the prototypes of complex concepts and those of their constituents. Finally, those theories are silent with regard to the type of properties that speakers regard as typical of each predicate in a given context. A fuller linguistic account ought to state the precise conditions under which a property counts as a feature.

3.5 How is semantic knowledge acquired?

How do speakers acquire their semantic knowledge about nouns' meanings and about their typicality ordering and feature-sets? This puzzle seems, prima facie, to be mainly the interest of experts in language acquisition. However, surprisingly perhaps, it appears that what we know about acquisition is directly related to typicality. I call the findings which support this observation *the order of learning effects*. They form some of the most robust typicality effects. In a nutshell, typical instances are acquired earlier than atypical ones, by children of various ages and by adults (Mervis and Rosch 1981; Rosch 1973); categories are learnt faster if initial exposure is to a typical member, than if initial exposure is to an atypical member, or even to the whole denotation in a random order (Mervis and Pani 1980); and finally, typical (or early acquired) instances are remembered best, and they affect future learning (encoding in memory) of entities and their features (Heit 1997). In sum, typicality is deeply related to the order in which instances are learnt to be members in predicate denotations.

Kamp and Partee 1995 do not explicate how the prototypes and measure functions – which their theory associates with predicates – might be acquired. In addition, the logical structure of a supermodel contains no intermediate stages in between M and each t in T (or under M). Thus, the order in which the denotations of each predicate grow in the process of language learning and the relations between this order and the typicality judgments cannot be represented. In contrast, in part 4, I propose that the graded typicality structure in each predicate (e.g., *bird*) stems from the order in which entities are learnt to be members in its denotations We encode this learning order in memory, either during acquisition, or even as adults, within a particular context, when we need to determine which birds a speaker is actually referring to (the contextually relevant set of birds). A considerable advantage of this approach is that while it is not quite clear which facts in the world the membership function in a supermodel actually represents, my proposal makes an explicit connection between the formal logical structure and some psychological entities (learning orders), whose existence is supported by highly robust and replicable empirical findings. I will now show how this proposal can be formally represented and how it solves the classical problems that we encountered.

4 My proposal: Learning models

Learning Models represent information growth, i.e., the order in which entities are categorized under predicates. We start with a zero context, c_0, where denotations are empty. From c_0 on, each context is followed by contexts in which more entities are added to the denotations. In total contexts, each entity is in the negative or positive denotation of each predicate. For example, birdhood is normally determined first for *robins* and *pigeons*, later on for *chickens* and *geese*, and last for *ostriches* and *penguins*. Similarly, non-birdhood is determined earlier for *cows* than for *bats* or *butterflies*:

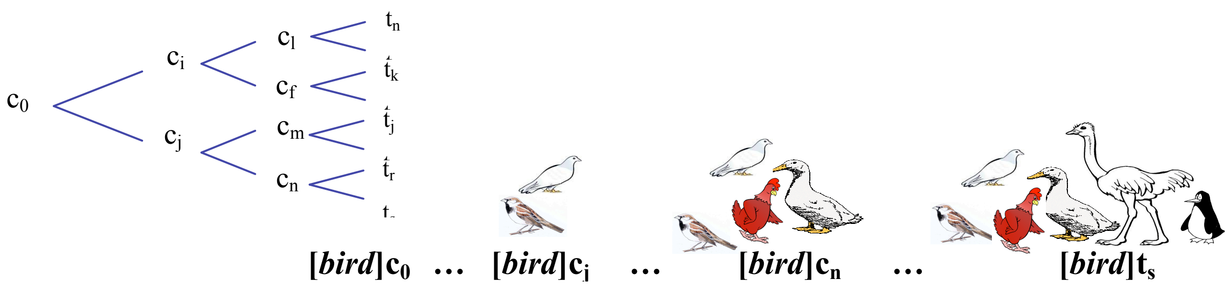

Figure 2: The contexts' structure in a Learning Model and an example of a branch in it

Formally, I use the model called *Data Semantics* (Veltman 1984; Landman 1991). A learning model M* for a set of predicates A and domain D is a tuple **$<C, \leq, c_0, T>$** s.t.:

(A) C is a set of partial contexts: In each c in C an n place predicate P is associated with partial positive and negative denotations (sets of n-tuples), **$<[P]^{+}_{c},[P]^{-}_{c}>$** .

(B) $\leq$ is a partial order on C: $\forall P \in A$:

1. **c_0** is the minimal element under $\leq$ (denotations are empty in c_0): $[P]^{+}_{c0}=[P]^{-}_{c0} = \varnothing$
2. Each t in **T** is maximal under $\leq$ (denotations are maximal in t): $[P]^{+}_{t} \cup [P]^{-}_{t} = D^{n}$
3. **Coherence**: The positive and negative denotations are disjointed: $[P]^{+}_{t} \cap [P]^{-}_{t} = \varnothing$
4. **Monotonicity**: $\forall c_1,c_2 \in C$, s.t. $c_1 \leq c_2$: $[P]^{+}_{c1} \subseteq [P]^{+}_{c2}$; $[P]^{-}_{c1} \subseteq [P]^{-}_{c2}$.

   (c_1 precedes c_2 iff predicate denotations in c_1 are subsets of the corresponding sets in c_2).
5. **Totality**: Every c has some maximal extension t: $\forall c \in C, \exists t \in T: c \leq t$
6. **"Super-denotations"**: $[P]_c = \cap\{[P]^{+}_{t} | t \in T, c \leq t\}$; $[\neg P]_c = \cap\{[P]^{-}_{t} | t \in T, c \leq t\}$

   The super-denotation of P in c contains, in addition to directly given Ps (i.e., members in $[P]^{+}_{c}$), also entities whose P-hood can be inferred (they are in $[P]^{+}_{t}$ in any t above c).

I propose that we consider d_1 *more typical of P* than d_2 in a context t iff under t, **either** the P-hood of d_1 is learnt before the P-hood of d_2 (i.e., in a context that precedes the one in which d_2 is added to the positive denotation) **or** the non-P-hood of d_2 is learnt before the non-P-hood of d_1 (in a context that precedes the one in which d_1 is added to the negative denotation):

(C) **$\forall t \in T$:** d_1 is *equally or more (typical of) P* than d_2 (**$<d_2,d_1> \in [\leq P]^{+}_{t}$**) iff:
$\forall c \leq t$: $(d_2 \in [P]_c \rightarrow d_1 \in [P]_c)$ & $(d_1 \in [\neg P]_c \rightarrow d_2 \in [\neg P]_c)$.
(in any context c under t, if d_1 is P, d_2 is P, and if d_2 is $\neg P$, d_1 is $\neg P$).
A *superdenotation* is defined, too: **$[\leq P]_c$** $= \cap\{[\leq P]^{+}_{t} | t \in T, t \geq c\}$.

5 The problems are solved

5.1 Intermediate typicality degrees for denotation members

In Part 5 we will see that a large number of long-standing puzzles are now solved. We will begin by noting that numerical degrees are not directly given. The primitive notion is of ordering, which is more intuitive. As for degrees, if needed, they can be derived from the ordering. An empirical study ought to tell precisely how (see more on this in 5.4). At any rate, vagueness with regard to degrees would be derived from gaps in the typicality ordering.

Now, recall that degrees of denotation members in Kamp and Partee's model are always maximal, i.e., 1. This is not the case in the current model. Rather, the earlier we learn that an entity is, e.g., a *bird*, the more typical we consider this entity to be. Crucially, although *bird* is a sharp noun, speakers encode different bird types in memory *gradually*. Moreover, the difference between vague and sharp predicates is not lost, but it becomes quantitative more than qualitative. We can say that the denotation of *bird* is already fully specified earlier in the context structure than the denotation of *chair*, which is more inherently vague.

5.2 Typicality in complex predicates

Another set of old problems (*subtype, conjunction* and *negation effects*) is also readily solved.

Subtype effects can now be accounted for. For example, the typicality degree of ostriches is predicted to be greater in the predicate *ostrich* than in *bird* iff they are categorized late in *bird*, relative to other bird types (robins, pigeons, etc.), but early in *ostrich*, relative to other ostriches! Since this is a natural state of affairs, in most contexts typical ostriches are indeed considered to be atypical *birds*. For example, in the birds model in Figure 2, the ostrich is *an atypical bird* in t_s. Yet, we can reasonably assume that this entity is the first member in the denotation of *ostrich* in t_s. Thus, it is *an atypical bird* and *a very typical ostrich* in t_s.

Conjunction effects are similarly accounted for. For example, the typicality degree of brown apples is predicted to be greater in *brown-apple* than in *apple,* whenever they are categorized late under *apple*, relative to other apple-types (*red*, *green*, etc.), but early under *brown apple*, relative to other brown apples. Similarly, the typical *male-nurses* are atypical *males* when the earliest known *males* are not *nurses*. They are atypical *nurses* when the earliest known *nurses* are not *males*. The above facts fall into place without the need for any new stipulations for complex predicates.

Furthermore, speakers, so it seems, do not use *typical* to refer to *likelihood* scales, but the other way around: They sometimes use *likely* to refer to typicality scales. If so, neither the conjunction effect, nor the so-called *conjunction fallacy,* are fallacies. Speakers' judgments may hold on to the status *rationale*. Indeed, the use of *typical* is normally constrained to utterances concerning the ordering relation that holds between denotation members. When the ordering relation that holds between gap members is concerned, the use of *likely* is preferred, and when the ordering relation that holds between non-members (members of the negative denotation) is concerned, the use of *similar to* or *related to* is preferred. These distributional constraints are not completely strict, but they do reflect general patterns of use.

Negation effects are also accounted for without any new stipulations. Given that in all contexts c, $[\neg P]^+_c=[P]^-_c$ and $[\neg P]^-_c=[P]^+_c$, the ordering of *non-bird* is, by the definition of typicality in (C), inverse to the ordering of *bird* in each context, (for supporting experimental evidence, see Smith et al 1988). But this inverse pattern is predicted only for the logical negation of a predicate. If a negated predicate like *non-bird* is contextually restricted to, say, *animals*, then it is not equivalent to the logical negation of *bird* and hence its ordering is not predicted to be inverse to the ordering of *bird*. In other words, in certain cases, entities are actually ordered by a conjunctive predicate, e.g., (*an animal which is*) *not a bird*, though the first conjunct (the one in the brackets) is covert. If so, it is only natural that certain entities, e.g., non animals, are assigned low degrees both in the predicate *bird* and in the latter conjunctive predicate: They belong to the negative denotation in both cases.

Additionally, a non-negated predicate (e.g., *bird*) might be contextually restricted, and again, this might affect its contextual ordering relation. For example, *chickens* usually precede *robins* in being regarded as both *birds AND walking in the barnyard*. Hence, their typicality degree in *bird* in a discussion about animals which might be *walking in the barnyard* is predicted to exceed that of robins, as Roth and Shoben 1983 indeed found. Thus, this type of context dependency of the typicality judgments can be reduced to typicality in conjunctions.

5.3 Partial and context-dependent knowledge

Yet another classical problem which learning models can handle well is the representation of partial knowledge about typicality. Typicality relations may be unknown: A pair, a penguin and an ostrich, say, is in the gap of the ordering *more typical of a bird* in a context c, iff it is still possible in c (i.e., true in some context following c) that the penguin is *more typical* in *bird*, and it is still possible in c that the ostrich is *more typical* in *bird*. Thus, the *inherent* context dependency of the typicality judgments can now be represented. No context (or valuation) independent ordering relations are part of the theory. Nor are any context-independent prototypes part of the theory. In each total context, *some* entities are the best in each predicate. In this way, we account for the ordering in, e.g., *typical tall person* despite the fact that there is no (context-independent) *maximal tallness*.

We also account for the fact that different ways to refer to ≤P differ in truth conditions. For instance, d_1 may be *more (of an) Italian* but *less typical (of an) Italian* than d_2 (if, say, d_1 lives in Italy but behaves like an American, while d_2 lives in America but still looks and behaves much like an Italian). We need not pose different definitional constraints on *more (of a) P*, *more typical (of a) P* and *more relevant (of a) P*. The difference between these three comparative types is pragmatic in nature. It is generally assumed that the comparative *more P* makes use of a semantic ordering dimension in the meaning of P (Kamp 1975). In contrast, *more typical P* makes use of different, or additional, ordering properties, namely, criteria from world knowledge, not just semantic criteria. Finally, *relevant P* makes use of additional, completely ad-hoc properties. These differences in the ordering criteria result in three diffrent ordering relations. For example, in *taller,* the ordering criterion (and hence relation) is fixed semantically, but in *more typical of a tall person*, we associate with *tall* more features (context-dependent ordering criteria). So the NP *typical tall person*, like *typical bird*, is associated with a context-dependent ordering relation. Luckily, since a different feature set and ordering relation may be associated with a predicate (e.g., *tall* or *bird*) in each total context in a learning model, this can be easily represented.

The elimination of the prototypes from the theory considerably simplifies the taxonomy of predicates: The distinction between predicates without a prototype, predicates with a prototype that does not affect the denotation, and predicates with a prototype that affects the denotation is eliminated. The intuitively-felt differences between these predicate types are accounted for, again, in a quantitative rather than qualitative manner. For example, the prototypical shade of *red* is universal and most likely innate or learnt very close to birth (i.e., early in the context structure), while the prototypical *non-bird* varies from context to context.

5.4 Learning models with typicality features

Finally, having stated what a typicality ordering is, the classical problem of defining the notion of *a typicality feature* (or an ordering dimension) and its roles can now be dealt with. Actually, we can adapt the assumptions of feature models and bypass their problems. The point is that, be the feature weights and the averaging function what they may, a *weighted mean hypothesis* will always entail that if a property like *flying*, or *being small*, counts as a feature of *bird* (i.e., for any non-zero weight for *flying*), any entity which is better in *flying*, and not worse in anything else, will have a higher average in *bird*. Therefore, as linguists, we do not need to discover the precise degree, weight and averaging functions which are being used in each context. A simple constraint may form an intuitive definition (or test) for *'feature'*-hood:

(D) *Ceteris-paribus correlation*:
A predicate F_1 counts as a typicality feature of a predicate P iff the ordering in F_1 correlates with the ordering in P *ceteris-paribus*, i.e., iff: Any entity more typical in F_1 than other entities, *and not less typical in other P features* (F_2, F_3, …), is more typical in P.

For example, a property like *flying* or *being small* counts as a feature of a predicate like *bird* iff any entity more typical in *flying* than other entities, and not less typical in other *bird*-features (e.g., *small*), is more typical of a *bird*. This is a *ceteris-paribus* (an all-other-things-being-equal) correlation. Exceptions (items which are more typical in *flying* but less typical in *bird* or vice versa) are allowed when and only when the ordering in two bird-features is inverse.

For a detailed discussion of a model with feature-sets, see Sassoon in progress. In a nutshell, in each context in the model, predicates are linked with both partial denotations ($<[P]^{+}_{c},[P]^{-}_{c}>$) and partial feature sets: $<F^{+}_{(P,c)},F^{-}_{(P,c)}>$. Feaures are themselves (simple or complex) predicates in the language. Take the predicate *chair*: Only in total contexts $t \in T$, any predicate (say – *wooden*) is either in $F^{+}_{(chair,t)}$ (*typical of a chair*) or $F^{-}_{(chair,t)}$ (*not typical of a chair*). If in a partial context $c \in C$, a predicate like *wooden* is neither in $F^{+}_{(chair,c)}$ nor in $F^{-}_{(chair,c)}$, then in c we do not yet know if being *wooden* is typical of *chair* or not. Now, c_2 extends ($\geq$) c_1 iff predicate denotations *and feature-sets* in c_2 are supersets of the corresponding sets in c_1. So this is a standard vagueness model, with feature sets.

Typicality features play the most basic roles in categorization processes. Constraint (D), together with the learning constraint (C), predict that knowledge of the *bird* features will trigger automatic categorization of new entities based on their features. For example, imagine that a chicken or a goose d_1 is already categorized as a *bird*. Then we encounter, say, – a duck d_2, and we can tell, based on constraint (D), that the duck is more typical in *bird* (because it is smaller, it flies better, etc.) Now, by constraint (C), if d_2 *is more typical than* d_1 and d_1 is a *bird*, then d_2 must be a bird. But imagine that we encounter an ostrich d_3. This time d_3 diverges, maybe too much, from the known birds. For instance, it is bigger than all of them. So, by constraint D, it is less typical, and by constraint (C), it is not necessarily a bird. These predictions are empirically supported. Kiran and Thompson 2003 found that, in aphasics and neural networks which are taught the category *features*, training with atypical members results in spontaneous recovery of categorization of untrained more typical items (like our duck), but not of untrained less typical items (like the ostrich). Similarly, in healthy subjects, previously inaccessible typical instances are frequently (falsely) assumed to be known once they become accessible (Reed 1988). Why? By constraint (C), if less typical entities are denotation members, entities which are more typical (given their higher average in the features) should already be known members, too! So both constraints (C) and (D) are at play.

But note that we have just demonstrated that sometimes inferences based on learning orders are suppressed. The learning constraint (C) predicts that in the duck situation, the duck would be considered *less typical* than the chicken (because it is learnt later to be a bird). But by virtue of constraint (D), we consider the duck *more typical* (because it is better in the bird features). Similarly, what if my initial exposure to birds was through ostriches?? We predict that based on constraint (C), I would think that ostriches are representative birds, and I would induce wrong category features, like wrong optimal size and *running* instead of *flying*. But not forever! Later on, I might decide to correct my beliefs, because, by constaint (D), ostriches cannot be representative. Their properties cannot be *bird* features, because besides the ostriches, the birds which are categorized earlier are *smaller*, they *fly*, etc. Again, constraint D seems to win. So, when does this happen? The answer is simple. Inferences based on constraint (C) are not replaced by inferences based on constraint (D), when and only when *all the entities under consideration become accessible to the language learner simultaneously*. Then, if *bird*-hood is taught earlier for some entities than others, this is a significant fact. But this was not the case in the given examples. Thus, if some entities are inaccessible in some contextual stages (say, the ducks), constraint D automatically corrects the distortions which emerge in the learning order. It forces us to 'jump' non-monotonically in the context structure, to a context in which, e.g., the ducks are in fact learned to be birds *earlier*, not *later*, than the geese and chickens, and the ostriches are categorized last.

Does this empty the learning hypothesis of its content? No. Constraint (C) is indispensible. It provides the learner with acquisition-based means to begin bootsrapping semantic knowledge about predicates' orderings. Constraint (D) only postulates connections between the comparative of a predicate and the comparatives of other predicates, its features. But this remains meaningless as long as these comparatives are not linked to a denotation – an entity ordering. However, once initial entity orderings are established, knowledge of features can be put to use and constraint (D) establishes its dominance.

We predict that learning (constraint C) affects the typicality judgments most in the lack of knowledge about the features. And indeed, experiments show that, in the lack of knowledge about features, first exposure to an atypical item slows down acquisition (Mervis and Pani 1980). This supports the proposal that such situations trigger the wrong inferences that the first given item is representative and that its features are typical of the category. These inferences are canceled later on when items that fall outside the category are discovered as averaging more highly in the inferred features than most category members. But this process is time-consuming and it slows down acquisition. Interestingly, in certain children, acquisition is completely blocked (within the experiment time). These children refuse to abandon the inferences based on constraint (C)! Furthermore, this is rarely acounted for in the literature, but *typicality* judgments also characterize proper names. For example, we can say that *smoking Nobles is typical of Fred*, meaning that any entity occurrence which is more typical in *smoking Nobles* than other entity-occurrences, *and is not worse in other Fred-characteristics,* is more typical of (an occurrence of) *Fred* (smoking). That is, one becomes more like a *Fred* if one smokes Nobles. Now, we are usually not taught the features of individuals we encounter in advance. And indeed, as constarint (C) predicts, when we determine the characteristics of individuals, usually the first impression that they make remains the most dominant.

To conclude, the proposed theory is more elegant (it stipulates no prototypes, degree functions, or modified degree functions) and yet it solves more problems, compared to Kamp and Partee 1995. It renders the 'muddy' psychological findings logical (not fallacious) and it improves our understanding of the ways typicality might affect utterance interpretation.

References

Barsalou, L.: 1983, Ad-hoc categories. *Memory and Cognition* **11:** 211-227.

Hampton, J.: 1997, Conceptual Combination. In: *Knowledge, Concepts and Categories*, K. Lamberts and D. Shanks (eds.), 135-162. Cambridge, MA: The MIT Press.

Heit, E.: 1997, Knowledge and concept learning. In: *Knowledge, Concepts and Categories*, K. Lamberts and D. Shanks (eds.), 135-162. Cambridge, MA: The MIT Press.

Kamp, H.: 1975, Two theories about Adjectives. In: E. Keenan (ed.), *Formal Semantics for Natural Language*.

Kamp, H. and Partee, B.: 1995, Prototype theory and compositionality. *Cognition* **57**: 129-91.

Kennedy, C. and McNally, L.: 2005. Scale structure, degree modification, and the semantics of. gradable predicates. *Language* **81(2)**: 345-381.

Kiran, S. and Thompson, C.: 2003, The role of semantic complexity in treatment of naming deficits: Training categories in fluent aphasia by controlling exemplar typicality. *Journal of Speech, Language and Hearing Research* **46**: 608-622.

Lakoff G., 1973, Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic* **2**: 458-508.

Landman, F.: 1991, *Structures for Semantics*. Dordrecht: Kluwer Academic Publishers.

Malt, B. and Sloman, S.: 2003, Linguistic diversity and object naming by non-native speakers of English, *Bilingualism: Language and Cognition* **6**: 47-67, Cambridge University Press.

Mervis, C. and Pani J.: 1980, Acquisition of basic object categories. *Cognitive Psychology* **12**: 496-522.

Mervis, C. and Rosch, E.: 1981, Categorization of natural objects. *Annual Review of Psychology* **32**: 89-115.

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

Reed, S.: 1988), *Cognition, Theory and Applications*. Pacific Grove: Brooks/Cole.

Rosch, E.: 1973, On the internal structure of perceptual and semantic categories. In: T.E. Moore (ed.), *Cognitive Development and the Acquisition of Language*, New York: Academic Press.

Roth, E. and Shoben, E.: 1983, The effect of context on the structure of categories. *Cognitive Psychology* **15**: 346-378.

Sassoon, G.: in progress, *Semantics with Clusters of Properties*. PhD diss., Tel Aviv Uni.

Smith, E., Osherson D., Rips, L, Keane, M.: 1988, Combining prototypes, a selective modification model. *Cognitive Science* **12**: 485-527.

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.

Sperber, D. and Wilson, D.: 1998, The mapping between the mental and the public lexicon. In: P. Carruthers and J. Boucher (eds.) *Language and Thought*. Cambridge University Press, Cambridge: 184-200.

Teigen, K. and Keren, G.: 2003, Surprises: low probabilities or high contrasts? *Cognition* **87(2)**: 55-71.

Tversky, A. and Kahneman, D.: 1983, Extensional versus intuitive reasoning: The conjunction fallacy in probability judgments. *Psychological Review* **90**: 293-315.

van Fraassen, B.C.: 1969, Presuppositions, Supervaluations and Free Logic. In: K. Lambert (Ed.), *The Logical Way of Doing Things*: 67-91. New Haven: Yale University Press.

Veltman, F.: 1984, Data Semantics. In: J. Groenendijk, T. Janssen and M. Stokhof (Eds.) *Truth, Interpretation and Information, Proceedings of the 3rd Amsterdam Colloquium*.

Wittgenstein, L.: 1968, Reprint: *Philosophical Investigations*, Translated by G.E.M. Anscombe, 3rd edition, Oxford: Blackwell. Original edition, 1953.

Zadeh, L.: 1965, Fuzzy sets, *Information and Control* **8**: 338-53.