Prototypes and their Composition from an Evolutionary Point of View

Gerhard Schurz

TPD PREPRINTS

Annual 2008 No.1

Edited by Gerhard Schurz and Markus Werning

Vorveröffentlichungsreihe des Lehrstuhls für Theoretische Philosophie an der Universität Düsseldorf
Prepublication Series of the Chair of Theoretical Philosophy at the University of Düsseldorf
Prototypes and their Composition from an Evolutionary Point of View

Gerhard Schurz (University of Duesseldorf, Germany)

Abstract: This paper applies an evolution-theoretic account of normic laws and prototypical properties to the prototype theory of concepts. After a brief outline of the evolution-theoretic account (sect. 1-2) and the prototype theory of concepts (sect. 3) I discuss five problems of prototype theory (sect. 4). These problems are dissolved by restricting the domain of prototype theory to evolutionary systems and by taking predictive and diagnostic efficiency as the main function of prototype information in conceptual content (sect. 5). Sections 6-9 deal with the question of compositionality. After distinguishing between formal and procedural compositionality (sect. 6) I argue that prototypes satisfy procedural compositionality only if the modifying adjectives do not express exceptional properties (sect. 7-8). I draw the conclusion that prototypes are semi-compositional (sect. 9).

1. Introduction: Normic Laws and Prototypical Properties

Normic laws are general conditionals of the form

\[ (1) \quad \text{As are normally Bs, here formalized as: } A \Rightarrow B. \]

(the terminology is due to Scriven 1959). They are not strict (i.e., strictly universal), but admit of exceptions (i.e., As which aren't Bs). Normic laws have been investigated in the areas of philosophy of science and of non-monotonic reasoning (cf. Schurz 2004). They have an obvious bearing to the

---

1 Work on this paper was supported by the Duesseldorf research group FOR600 funded by the DFG. For various help I am indebted to Markus Werning and two anonymous referees.
prototype theory of concepts, insofar as the normic law (1) can be equivalently rephrased as follows

\[(1^*) \text{ B is a prototypical property of As (in the wide sense of 'prototypical').}\]

A more narrow sense of 'prototypical properties' will be introduced in section 2.

In the 1950s, philosophers of science have discovered that there are almost no strict laws to be found in the 'higher' non-physical sciences. Rather, the typical explanations in these disciplines make use of normic laws, such as the following:

(2) People's actions are normally goal-oriented (folk psychology).
(3) Birds normally can fly (biology).
(4) Turning the ignition key normally turns on the engine of my car (technology).

In this and the next section I will develop an evolution-theoretic explanation of the omnipresence of normic laws and prototypical properties in everyday life and in non-physical sciences. My explanation consists of two parts. First, I will argue that reasoning with normic laws is efficient and reliable. Second, I will show that human's environment is populated by evolutionary systems which obey normic laws, which explains why human cognition is well adapted to prototypicality structures.

In order to enable reliable reasoning, the concept of (prototypical) normality must have an objective meaning which connects it with statistical probability, in the sense that if B is a prototypical property of As, then most As are Bs. The thesis that normic laws entail a numerically unspecific statistical majority claim is henceforth called the:
Statistical consequence thesis (SC): $A \Rightarrow B$ implies that the conditional statistical probability of $B$ given $A$, $P(B|A)$, is high.

The statistical consequence thesis does not imply that normic laws can be replaced by numerical conditional probabilities. The numerical probabilities associated with prototypical properties are typically highly context-sensitive, so that we are unable to specify them without specifying the context. For example, how many percent of all birds can fly (when?, where?). All that one knows or assume – according to the statistical consequence thesis – is that these probabilities are high.

Early philosophers of science as well as founders of non-monotonic reasoning have doubted that normic laws have objective meaning – they suggested that normic laws would be similar to linguistic conventions (cf. Dray 157: 132; Scriven 1959: 466; McCarthy 1986; Reiter 1987). The crucial problem of this position is, of course, that if normic laws are taken as conventions, then nothing guarantees that reasoning from prototypical properties are reliable, i.e. leads to true predictions in a high majority of cases. In other words, the reliability of normic laws requires an objective characterization of prototypical properties which connects them with statistical reliability (cf. Hempel 1965; Pearl 1988: 477-80).² But the question is – why should such a connection hold? Why should the objective probability distributions over the properties of complex 'living' systems possess so many high (conditional) probability peaks, rather than being more or less flat? As long as we do not have an explanation of the statistical consequence thesis, the force of those doubts will remain strong who maintain that the omnipresence of normic laws, instead of being the result of objective feature of reality, is merely the product of our subjective framing of a world whose complexity exceeds our cognitive limitations.

In the next section I will present such an explanation which has been developed in Schurz

² An alternative proposal to 'objectify' normic laws is to reconstruct them as ceteris paribus laws. Various problems of this account are discussed in Schurz (2002).
(2001a). It is based on an evolution-theoretic understanding of prototypical properties as the result of evolutionary adaptations which establish a connection between prototypicality and high statistical probability on nomological reasons. This explanation has the following advantages:

1) it makes the statistical consequence thesis plausible,

2) it is needed to defend the statistical consequence thesis against objections of philosophers of biology (e.g. Millikan 1984, Laurier 1994, Wachbroit 1994), and

3) based on 1) and 2), it yields a deeper understanding of the achievements as well as the limitations of prototype semantics.

2. Generalized Evolution-Theory as a Foundation of Normic Laws and Prototypical Properties

All 'higher' sciences, from biology upwards, are concerned with living systems or with their cultural and technical products. What these systems have in common is the characteristic capacity of self-regulation under the permanent pressure of their environment. So this is my first thesis:

Thesis E1: Normic laws describe the properties of self-regulatory systems. According to the framework of cybernetics (cf. Ashby 1961), the identity of self-regulatory systems is governed by certain prototypical norm states (or properties) which these systems constantly try to achieve and maintain with the help of regulatory mechanisms which compensate for disturbing influences of the environment.

But what explains the omnipresence of self-regulatory systems in our world? And what explains their proper functioning most of the time? The answer is contained in my next thesis:
Almost all self-regulatory systems are evolutionary systems in the generalized 'Darwinian' sense. They have evolved through a recursive (natural or cultural/technical) process of reproduction, variation and selection. Their prototypical norm states and self-regulatory mechanisms have been gradually selected in their evolution history according to their contribution to reproductive success.

Evolution theory explains why evolutionary systems obey normic laws which imply high conditional statistical probabilities. The self-regulatory capacities of evolutionary systems are limited. Dysfunctions may occur, hence their normic behaviour may have various exceptions. Yet it must be the case that these systems are in their prototypical norm states most of their times – for otherwise, they would not have survived in evolution. Green plants, for example, can normally perform photosynthesis. Of course it is possible that due to a catastrophic event, all green plants lose this ability. But then (with high probability), they will become extinct after a short period of evolution. For similar reasons, electric devices normally work, for they are constructed in that way, and if this were not so, they could not survive in the economic market. Put in a nutshell, prototypical normality and statistical normality are connected by the law of evolutionary selection.

Various refinements of this rather crude presentation of my evolution-theoretic foundation of normic laws can be found in Schurz (2001a). For example, to be applicable to normic laws of all higher sciences, the account has to be based on the generalized theory of evolution in the sense of Dawkins (1989, ch. 11). Generalized evolution theory does not reduce evolution to the evolution of genes as the reprotypes of biological evolution, but assumes in addition an independent level of so-called memes (acquired cognitive abilities) as the reprotypes cultural evolution. The reprotypes which underly a kind of evolution are those entities which get directly reproduced, while the so-called phenotypes are the 'prototypical' properties which are produced by the reprotypes in the given
environments and which are the target of selective forces. The following three points illustrate how my evolutionary account can explain normic laws without entailing excessive adaptationism:

1.) Whenever an evolutionary system (species) $S$ contains two competing (heritable) traits $P$, $P'$ such that the $P$-variants have a higher reproductivity (‘fitness’) than the $P'$-variants, then after sufficiently many generations the frequencies will be driven to almost 100% for $P$-variants and almost 0% for $P'$-variants, independently of the initial frequencies of $P$-variants and $P'$-variants. In this way, the normic law "Ss are normally Ps" originates from standard evolutionary selection conditions.

But note that selectively advantageous properties are never driven to strict fixation: a small remnant of 'abnormal' variants is constantly produced by spontaneous variations. This small percentage of 'abnormal' species members is not superfluous but highly important for the persistence of the species under sudden changes of environmental fitness conditions (cf. Ridley 1993: 204ff).

2.) Not all properties are positively or negatively prototypical, i.e. are either possessed by the majority or by the minority of the species members, on two reasons: (2.1) Only biologically or culturally reproduced (herited) properties can count as prototypical; properties whose existence depend on contingent constellations of the environment can't. (2.2) Among the reproduced properties there may exists cases of so-called polymorphisms with arbitrary (e.g. uniform) probability distributions. However, such polymorphisms occur only under non-standard selection conditions (such as multiple niche polymorphism, heterozygotic superiority, or negative frequency-dependence; cf. Ridley 1993, ch. 2). In standard evolutionary scenarios, property distributions will acquire prototypicality structure as explained in 1.) above.

3.) Not all prototypical properties are functional in the sense of having conferred a selective advantage To take a much debated example (Bigelow and Pargeter 1987), circulating the blood is a proper function of the vertebrate heart, while the typical sound of the heart beat is a mere side effect of it – but still it is prototypical for vertebrates' hearts to make this sound. A subtle example of this

---

3 An excellent overview of generalized evolution theory is found in Mesoudi et al. (2006).
sort are Gould and Lewontin's 'spandrels' (1979), which are prototypical side-effects of complex structural-anatomical architectures which were selected for independent reasons. To cover this difference, I distinguish between fundamental versus derived prototypical traits (see def. (P) below): while the former confer a direct selective advantage, the latter are mere causal side-effects of the former.

In Schurz (2001a) I arrived at the following definition of prototypical normality in the evolution-theoretic sense:

**Definition (P):** (1.) P is a prototypical property of members of a class of evolutionary systems S i.w.s. (in the wide sense) iff P is produced by a reprotype R and in the evolutionary history of S there was overwhelming selection in favour of R.

(2.) If the selection mentioned in (1.) took place because of R's producing P, P is called a fundamental prototypical property, and otherwise (i.e., if P is just a side-effect of R) P is called a derived prototypical property.

Millikan's concept of an (evolutionary) proper function (1984, ch.1) can be derived from this definition as follows: if a fundamental prototypical trait of S-members consists in the possession of item (organ) X with effect F, then F is called a proper function of item X of S-members. The major objection against the statistical consequence thesis given by Millikan (e.g. 1984: 4f, 34) and Laurier (1994: 29-31) points out that many items perform their functions not frequently but just often enough to get selected. For example, living up to the reproductive age is the major proper function of biological organisms, although only a small minority of babies of most species stay alive. This objection is resolved as follows: prototypical property do not consist in the actual performance of a function. Actual performances – for example, whether a baby survives or not – depend on accidental circumstances; they are not reproduced (herited) and, hence, cannot be claimed to be prototypi-
cal. What is prototypical is rather the *capacity* to perform the functional behaviour under certain triggering circumstances (which themselves need not be frequent). These capacities are the fundamental prototypical properties which are reproduced and selected as the result of the genetic (or memetic) constitution of S-members; with the result that under standard selection scenarios their frequencies will be driven to almost-fixation.

In Schurz (2001a: 494f) I gave a proof that the statistical consequence thesis of sect. 1 is a logical consequence of definition (P) together with mild assumption about evolution. Let me illustrate definition (P) by way of some examples. It is a prototypical capacity of matches to light when struck because they have been selected for this effect. It is a prototypical side-effect of matches that their flame sometimes burns one's fingers. But it is not prototypical for matches to have a certain colour, etc. It is a fundamental prototypical trait of human noses to smell and to stick out from the face. It is also prototypically normal for human noses to get cold at their top in the winter, but merely in the derived sense. Having legs, clearly, is prototypically normal for humans (in the fundamental biological sense), but not having short or long longs, because there was no dominant biological selection for short versus long legs.

Prototypical normality applies not only to 'species' but also to higher order classes of evolutionary systems. For example, flying-ability is (fundamentally) prototypically normal within the entire class of birds, although for certain (exceptional) species of birds, such as emus or penguins, lack of flying-ability is prototypically normal. This is not a logical conflict, but just illustrates the *non-monotonicity* of prototypical properties – they admit exceptions, *not* only at the level of abnormal individuals (a bird with defective wings) but also at the level of abnormal subspecies whose selection conditions have been significantly different from those of the super-species S. Penguins, for example, have changed their ecological niche from air-living to water-living creatures with the evolutionary effect that their wings have been gradually reshaped by selection into swimming instead of flying organs (a process which is called ex-adaptation).
One may ask, why do we then categorize penguins still as birds? Given that we base our categorization on prototypical properties, then the answer is that penguins still have sufficiently many prototypical properties in common with birds to be classified as an (exceptional) bird-species. One may also ask, what would happen if in some distant future the frequency of birds without flying ability would increase up to a value of 50%? In that distant future, flying ability could not longer be called a prototypical property of birds, because there is no longer overwhelmingly positive selection in favour of flying ability among birds. This point highlights the *historical* character of the prototypicality in evolution. The historical character is more salient in cultural evolution which proceeds much more rapidly – for example, the prototypical properties of farms two centuries ago (with horses, ploughs, etc.) are completely different from contemporary farms.

The relation between super- and subclasses of evolutionary systems is connected with the important distinction between prototypical properties in the *wide* sense (i.w.s.) and in the *narrow* sense (i.n.s.). Definition (P) identifies prototypical properties in the *wide* sense as those properties of an evolutionary system S which were selectively advantageous for the evolutionary persistence of S. While some of these properties are *indicative* for S, some others are shared by S with many evolutionary 'sibling' species of S which derive from a common ancestor species. The prototypical properties of S i.n.s. (*in the narrow sense*) are *by definition* those prototypical properties of S i.w.s. which are indicative for S, i.e. which are *not* also prototypical properties of evolutionary sibling species of S. For example, having wings is a prototypical property i.n.s. of birds, but it is merely prototypical i.w.s. (and not i.n.s.) for sparrows or other special kind of birds, because it does not discriminate sparrows from other kinds of birds. Likewise, tasting sweet and being juicy is a prototypical property i.n.s. of fruit which is also a prototypical property i.w.s. but not i.n.s. of ripe apples, because most other kinds of ripe fruit have this property, too. These examples show that the use of 'prototypical properties' is systematically ambiguous between i.w.s. and i.n.s. – whence I suggest to distinguish between these two senses. If one emphasizes the *discriminative* value of prototypical
properties (cf. Rosch and Mervis 1975; Kleiber 1998: 48ff), one means the narrow sense, while if one assumes the default inheritance of prototypical properties from super-categories to sub-categories (cf. Collonny et al. 2007) then one means the wide sense of prototypical properties.

The overall picture of my evolution-theoretic account is this: our natural, cultural or technical environment is full of different kinds evolutionary systems, organisms and artifacts, each of them being characterized by a list of prototypical properties which have played a decisive role in their selection history. Given that humans' environment is full evolutionary systems whose behavior is governed by prototypical properties, it should be expected that human reasoning is fit in reasoning with prototypes. In (Schurz 2005) I set up two criteria of fitness of human reasoning with prototypes: normic (prototype) reasoning must be (i) statistically reliable and (ii) feasible by algorithms of low complexity. I tried to show that the inference rules of the system \( P \) of probabilistic default reasoning (which are described in § 8) satisfies both criteria. Moreover, in several psychological experiments it has been confirmed that the basic inference rules of the system \( P \) are well entrenched in people's intuitive reasoning (cf. Schurz 2007; Pfeifer & Kleiter 2005: 2008; Evans et al. 2003). In the next sections I extend this line of argument: given my description of evolutionary systems is correct, it should also be expected that the concepts by which humans describe evolutionary systems reflect their prototypicality structure. This brings me to the evolution-theoretic account of the prototype theory of concepts, which is described in the next sections.

3. The Prototype Theory of Concepts

The prototype theory of concepts was developed in the 1970s by E. Rosch in order to overcome certain difficulties of the classical theory of the meaning. Several versions of the prototype theory have been developed since then (for an overview cf. Kleiber 1998, Margolis and Laurence 1999), and they have the following in common:
1.) The prototype theory refutes the classical theory of concepts according to which the meaning of a concept (or extensionally: a category) is defined by a list of individually necessary and jointly sufficient properties or conditions which an individual must possess in order to instantiate that concept (to be a member of the corresponding category). In fact, most of everyday language concepts do not have definitions by necessary and sufficient conditions. Nobody has ever given a definition of a bird, a dog, or a chair. All that can be given are prototypical properties of birds, dogs or chairs which are not necessary but allow of exceptional instantiations. For example, prototypical birds can fly, but penguins cannot; prototypical chairs have four legs, but some chairs have three legs, etc. The prototype theory of concepts identifies the meaning of a concepts with a certain prototype – which is either a prototypical exemplar of the category, or a list (or structure) of prototypical properties of the category.

2.) The prototype theory differs from the theory-theory of concepts according to which the meaning of a concept is provided by a given background theory. The theory-theory is adequate for two kinds of domains: (i) theoretical concepts of scientific theories, and (ii) certain anthropologically universal common sense concepts such 'inanimate' versus 'animate' being, 'cause' and 'effect', 'person', 'animal', 'plant', etc., which seem to be provided by inborn theories (cf. Margolis 1999: 43ff; Spelke 1990, Carey 1985). However, for the majority of everyday language concepts such as 'bird', 'chair' or 'fruit', common sense lacks background theories of this sort.

Generally speaking I support a pluralistic meaning theory: the meaning of some concepts is given by explicit definitions, that of others it is given by background theories – and prototype theory seems to be the right semantic approach for those concepts whose meaning is given by common sense experience. Various experimental evidence supports the mental reality of prototypes. For example, the ratings of the degree of typicality of exemplars or subkinds of a concept – e.g. the typicality of a sparrow or a pelican as representative for a bird – are surprisingly coherent among different test persons. Moreover these typicality ratings are positively correlated to the speed and nega-
tively to the error rate with which test persons classify given exemplars as member of the respective category (cf. Smith and Medin 1981; Osherson and Smith 1981: 263; Smith et al. 1988: 356; Kleiber 1998: 32). Prototype theory is also supported by recent semantical theories which stress the perceptual content of concepts, since memorized visual images focus one prototypical properties (cf. Barsalou 1999, Prinz 2002).

In earlier versions of prototype theory (e.g. Heider (= Rosch) 1971, Rosch 1975) the prototype was understood as a typical exemplar or subkind of the given concept. For example, a 'sparrow' is frequently judged to be the prototype of a bird, or 'apple' to be the prototype of a fruit. The degree of typicality, in short $d^{\text{typ}}(x:C)$, in which an exemplar $x$ is an instance of the concept $C$, was identified with the degree of membership, i.e. with a fuzzy measure of the degree in which $x$ is a member of $C$'s extension. However, in a seminal paper Osherson and Smith (1981) have demonstrated that the distance function of fuzzy set theory does not adequately model prototypes. The major obstacle is the so-called conjunction effect: an individual $x$ (e.g. a given brown apple) may be a bad member of a super-ordinate category $A$ (for apple) but still a good member of a non-prototypical subcategory $B&A$ (for brown apple) (cf. Hampton 1982; Smith and Osherson 1984; Smith et al. 1988: 357). In this case, $d^{\text{typ}}(x:A&B)$ would have to be greater than $d^{\text{typ}}(x:A)$, which is impossible for fuzzy set theory, since according to the min-rule of fuzzy set theory, $d^{\text{typ}}(x:A&B) = \min\{d^{\text{typ}}(x:A), d^{\text{typ}}(x:B)\}$.

In reaction to this problem, other authors have suggested to understand the prototype of a concept as a list of its prototypical properties. An elaborated model of this kind has been developed and experimentally confirmed by Smith et al. (1988). These authors represent a prototype by a prototype frame. Frames decompose properties (such as 'red', 'green') into attributes (e.g. colour) and values (red, green …); they represent concepts by lists of attributes together with their admitted values. Prototype frames assign, in addition, to each admitted value of an attribute its number of

---

4 Also Rosch (1978) had modified her original position.
'votes' which measures the typicality of this value for the prototype and is correlated with the frequency in which the value occurs among perceived instances (Smith et al. 1988: 358). Moreover, to each attribute a second number is assigned, the so-called diagnosticity of this attribute, which measures the usefulness of this attribute in discriminating instances of the concept from instances of contrasting concepts (the importance of diagnosticity has been experimentally confirmed by Rosch and Mervis 1975).

In the remainder sections of this paper I will focus on prototype frames because they are a rather advanced version of prototype theory which is especially suited for my purposes, in particular because their votes are connected with the statistical frequencies of the attribute-values of evolutionary systems. For example, the prototype frame of 'apple' is modelled by Smith et al. (1988: 358) as follows:

<table>
<thead>
<tr>
<th>Attribute (diagnosticity)</th>
<th>Values (votes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colour (1)</td>
<td>Red (25)</td>
</tr>
<tr>
<td></td>
<td>Green (5)</td>
</tr>
<tr>
<td></td>
<td>Brown (0)</td>
</tr>
<tr>
<td>Shape (0.5)</td>
<td>Round (15)</td>
</tr>
<tr>
<td></td>
<td>Square (0)</td>
</tr>
<tr>
<td></td>
<td>Cylindrical (5)</td>
</tr>
<tr>
<td>Texture (0.25)</td>
<td>Smooth (25)</td>
</tr>
<tr>
<td></td>
<td>Rough (5)</td>
</tr>
<tr>
<td></td>
<td>Bumpy (0)</td>
</tr>
</tbody>
</table>

Further attributes may be taste (sweet 20, sour 5), consistency (juicy 20, dry 5), etc. The degree of typicality of a given instance I is computed as

\[
\text{d}^{\text{typ}}(I: \text{Apple}) = \sum_{i \in \text{Att}} \delta_i (v_i(P \cap I) - v_i(P - I) - v_i(I-P)),
\]

where \(\delta_i\) is the diagnosticity of attribute \(i\), \(v_i(P \cap I)\) is the number of votes for values of attribute \(i\).
which the prototype P and the instance I have in common, \(v_i(P-I)\) is the number of votes for values of attribute i which are unique for P, and likewise for \(v_i(I-P)\). For example, if I is a particular apple with 30 votes for red, 20 for round and 30 for smooth, while votes for all other values of attributes are zero, then \(d^{pp}(I:Apple) = 1 \cdot (25-5-5) + 0.5 \cdot (15-5-5) + 0.25 \cdot (25-5-5) = 15 + 2.5 + 3.75 = 21.25\) (see Smith et al. 1988: 358f).

4. Five Objections against Prototype Theory

In this section I present five objections against prototype theory which have been discussed in the literature:

4.1 The problem of analyticity and intersubjective meaning stability. Fodor (1984: 26f) has defended the distinction between semantic (analytic) and empirical (synthetic) information against the attack of Quine (1951) by pointing out that a central function of this demarcation is to enable reasonable communication between speakers having different background beliefs. For if radical meaning holism were true then two speakers with different opinions about say priests would mean different things when talking about priests. To be sure, neither Fodor nor myself support a traditional (Frege-Carnapian) view of meaning postulates – all what I need to assume is that every adequate theory of meaning has to assume some sort of semantic-factual-demarcation.

Now, the prototypical frame of a concept does not only contain purely semantic information; it also contains world-knowledge, i.e. factual information (cf. Kleiber 1998: 48-54). For example, it is an empirical information that most apples are yellow and red, but unripe apples are green. But the empirical background knowledge of different people is different. So the question arises how the meanings of concepts as represented by prototypes can be intersubjectively stable? According to Wierzbicka (1985: 40f, 115), the prototype of a concept should not contain expert knowledge but only common sense knowledge. But the problem of intersubjective (in)stability arises also for
common sense knowledge. As an example, consider the process of concept acquisition in the childhood: some children may learn the prototype of a bird at hand of sparrows, because there are sparrows in their vicinity, while other children may learn the bird prototype at hand of pigeons or crows, etc. How does it come that, as a matter of fact, subjects agree in their concept of bird to such a high degree, in spite of the differences in their history of learning the concepts via prototypical exemplars?

4.2 Typicality versus Vague Membership: Many authors have pointed out that degrees of typicality have to be sharply distinguished from degrees of vague membership: $d^{typ}(x:C) \neq d^{memb}(x:C)$. Only vague membership leads to a fuzzification of the categorization problem, insofar it entails a fuzzification of the membership relation, while degrees of typicalities are compatible with strict categorization. For example, ducks, vultures or pelicans are rather untypical birds; but they are no less birds than sparrows (cf. Kamp and Partee 1995: 133f; Lakoff 1986: 43; 1987; Kleiber 198: 106ff.). While for Kamp and Partee (1995: 169), $d^{typ}(x:C) = d^{memb}(x:C)$ holds at least for some concepts, e.g. for 'chair' or for 'red', Osherson and Smith (1997: 192) argue convincingly that even for the latter concepts, $d^{typ}(x:C)$ and $d^{memb}(x:C)$ diverge: for example, a 'chair made of gold' is untypical but still clearly a chair, and according to Berlin and Kay (1969, Appendix I: 119), the wave-length spectrum of 'unambiguously red' is much more comprehensive than that of 'prototypical red'.

4.3 What is the semantic (or cognitive) function of prototypes? If it is not the function of prototypes to provide criteria for categorization (i.e. concept-membership), what is their semantic or cognitive function at all? Some authors have argued in favor of a dual theory of meaning, according to which each concept possesses a certain core meaning, which is purely semantic (or analytic), and a certain portion of (synthetic) world-knowledge about prototypical properties.\(^5\) The core meaning is

---

\(^5\) Cf. Osherson and Smith (1981, 277), Landau (1982). The position according to which there don't exist sufficient but at least necessary semantic conditions for membership is called the neo-classical theory (cf. Margolis and Laurence 1999, 52ff.).
compositional (see below) and provides \textit{necessary} conditions for the membership-relation. For example, whatever belongs to the category 'lion' \textit{must} be an animal; 'stone lions', although they have several prototypical properties in common with true lions, are \textit{not} lions. The semantic core meaning is also responsible for another important task which according to Fodor (1999: 176) a semantic theory should fulfill, namely the specification of satisfaction conditions for sentences: these are derivable from the conditions for membership relations together with the meaning of logical operators.

Even if one accepts the dual theory of meaning, the question concerning the semantic or cognitive function of the second meaning component, the prototype, remains. For Lakoff (1987a), prototypes are a mere 'side product' of categories whose members bear a \textit{family resemblance} to each other (cf. Kleiber 1998: 113). In any case, the question of the precise function of prototypes becomes pressing. Note that this problem has two sides: (a) what is the \textit{cognitive} function of prototypes (if any), and (b) can this cognitive function be regarded as a \textit{semantic} function?

4.4 Concepts without prototypes. Many concepts do not possess prototypes at all (cf. Margolis and Laurence 1999: 36, 44). For example, the following two kinds of concepts do not possess prototypes because their range of instances is too \textit{heterogeneous}: (1) Metric concepts: e.g., there is no prototype of objects which are longer than one meter, or hotter than 100 degree Celsius; and (2) \textit{negative} or \textit{disjunctive concepts} such as 'being not a bird', or 'being a bird or a table'. Other kinds of concepts do not possess prototypes because they are (3) \textit{too abstract} (e.g., 'dimension'), or they are (4) \textit{too theoretical} (e.g., 'mass' and 'force' in classical physics). Even some primitive kind concepts may become so heterogeneous in linguistic evolution that they \textit{loose} their common prototype. An example of this sort is Lakoff's \textit{mother}-example (cf. 1987b: 400): there is no common prototype underlying all present linguistic uses of 'mother' – such as genetic mother, birth mother (giving birth to the child, nowadays not necessarily identical with the genetic mother), foster mother, adoptive mother or stepmother. All these counterexample imply that the prototype theory of concepts has a \textit{restricted} domain of applications. How can we describe these restrictions in a \textit{non ad-hoc} way?
4.5 Major Problem: The Non-Compositionality of Prototypes. Fodor and Lepore (1996) have fundamentally criticized the prototype theory because of its apparent non-compositionality. The principle of compositionality requires that the meaning of a complex expression is fully determined by the meaning of its primitive syntactic constituents together with the (semantic composition functions underlying the) syntactic rules by which it is composed from these constituents. In many examples, however, the prototype of a complex expression does not appear to be determined by the prototypes of its constituents. In Fodor and Lepore's famous example, the prototype of a 'pet fish' (for example, a gold fish) is neither prototypical for a pet nor prototypical for a fish. Prototype theorists have tried to rescue compositionality for prototypes by modelling the semantic composition function as a more complex operation than mere conjunction – but it is still hard to see how any purely semantic composition function could succeed in retrieving the prototypical properties of 'pet fish' from the prototypical properties of 'pet' and 'fish' (more on this in §7).

The requirement of compositionality is usually justified by the argument that any theory of meaning must be able to explain the following two fundamental features of natural language (cf. Fodor and Lepore 1996: 28; Robbinns 2002: 315; Connolly et al. 2007: 2):

(i) productivity: competent speaker can understand indefinitely many complex expressions based on finitely many primitive meaningful terms and primitive syntactic operations, and

(ii) systematicity: what a competent speaker understands is closed under forming new combinations (e.g. 'red chair') from the constituents of already understood expressions (e.g. 'red table' and 'white chair').

Defenders of compositionality argue that every adequate meaning theory must satisfy composi-

---

6 In the literature, the principle of compositionality is usually stated without the brackets (e.g. Fodor and Lepore 1998, 29, fn. 3; Robbins 2002, 314; Connolly et al. 2007, 3), but in more technical papers it is stated with the refinement in brackets (cf. Hodges 2001). Assuming that the relation between syntactic rules and their corresponding semantic composition functions is one-to-one, both versions are equivalent.
tionality because compositionality is the only or at least the best explanation of productivity and systematicity. Margolis and Laurence (1999: 42) assert that the criticism of non-compositionality has no force against the dual theory of meaning explained in § 4.3. However, the problem cannot be solved so easily: although according to the dual theory the analytic core meaning of expressions is still compositional, the extended prototype meaning is not compositional. So the dual theorist would have to conclude that productivity and systematicity holds only in a restricted form: we can retrieve the meaning of newly heard complex expressions from their well-understood constituents only partially but not fully.

5. Evolution-theoretic Foundation of Prototype Theory

In spite of the objections discussed in sect. 4, prototype semantics is nevertheless adequate if its domain is restricted to common sense concepts describing evolutionary systems, and if its function is considered from an evolutionary viewpoint. In other words, the purpose of this paper is not a universal defense of prototype theory, but the attempt to give prototype theory its natural place in a pluralistic theory of meaning. The consequences of my evolution-theoretic account (recall sect. 2) for prototype semantics can be summarized in the following two theses:

**Thesis P1:** One (if not the) major evolutionary function of cognition is efficient predictive reasoning (inferring the effects of practically important causes) and diagnostic reasoning (inferring the causes of practically important effects). Categorization is a necessary condition of predictive and diagnostic reasoning, but categorization per se is not evolutionary advantageous because not every categorization is predictively and diagnostically efficient.

Thesis P1 contrasts with those cognitive scientists who consider categorization per se as the major
task of cognition (cf. Lakoff 1987a, XI; Kleiber 1998: 4f). For example, the manifold systems of mythical categorizations in animistic and early religious world-views are neither predictively nor diagnostically efficient. The category 'bayi' of Australasian aboriginees which covers "man, cangarooh, bats, most snakes, … boomerang, certain spears …" (Kleiber 1998: 122) may be fascinating for anthropologists but is predictively hopelessly inefficient; categories of that sort can hardly survive in the evolution of cognition. Systems of categorization are predictively and diagnostically efficient if they possess computationally simple categories which figure as junctions in a dense system of lawlike connections (cf. also Rosch 1978).

---

**Thesis P2:** The proper domain of prototype theory are evolutionary systems (as described in § 2). In this domain predictive and diagnostic efficiency is achieved because (following from sect. 2) evolutionary systems obey the following principles:

- **P2.1** Each species (or kind) S of an evolutionary system is characterized by a bundle of prototypical properties i.w.s. which have been selected during the evolution of (ancestors of) S and which S-members possess with a high statistical frequency.

- **P2.2** A certain subset of these prototypical properties i.w.s. – namely the prototypical properties i.n.s. – are highly discriminative vis-a-vis sibling species of S, because each kind of evolutionary system had its specific adaptation history to the selection requirements of its environment.

- **P2.3** All kinds of evolutionary systems include exceptional exemplars or subkinds, which deviate from the prototype pattern of the kind.

---

Theses P2.1-2 entail that the statistical distribution function over the total multidimensional

---

7 Some aspects of mythical and religious categorizations can be explained by their 'generalized placebo effect' (Taylor 1989, Schurz 1998) or by their effects on promoting altruistic behavior (Wilson 2002).
property space will have several sharp peaks corresponding to evolutionary kinds, with low and broad valleys between these peaks. In other words, only certain combinations of properties have been produced by evolution and not arbitrary combinatorial variations in between them. There are no dogs with wings or birds with teeth, etc. This is the reason why reasoning with prototypes is so fast and predictively so efficient, although by thesis P2.3 the laws of evolutionary systems are never strict but merely high majority laws. One prototypical property i.n.s is usually sufficient to identify an individual as a member of an evolutionary (natural, cultural) kind with high statistical reliability, and from this identification we can predict instantly a variety of other prototypical properties i.n.s. or i.w.s. Since we use all information we can have, every prototypical property i.n.s. can be used as means of identifying an evolutionary kind. For example, members of a primitive tribe hear a prototypical sound, or see a prototypical foot print, and instantly predict there is a predator nearby. Or, they see a prototypical shape of landscape and instantly predict there is a river (etc.).

A further clarification is important: since human prototype reasoning is more or less unconscious (Schurz 2007: 629), it may of course happen that humans apply prototype reasoning also to entities which are not evolutionary systems – e.g. they even speak of prototypical 'prime numbers' (cf. Armstrong et al. 1983) – but whenever they do this prototype semantics and prototype reasoning looses its efficiency.

Some authors argue that the function of prototypes does not consist in categorization per se, but in (heuristic rules for) fast categorization (cf. Machery in this volume). This is true, but it is just one half of my story. To enable efficient predictions, prototypes must enable fast categorization – i.e., the inference from a prototypical property P\textsubscript{i} i.n.s to the category C (P\textsubscript{i} \Rightarrow C) must be computationally easy and reliable – but in addition, categories must figure as junctions in a network of correlated prototypical properties, i.e. the inference from C to many (further) prototypical properties i.w.s. (C \Rightarrow P\textsubscript{j}, for j\neq i) must be computationally easy and reliable, too. From the two inferences P\textsubscript{i} \Rightarrow C and C \Rightarrow P\textsubscript{j}, the direct predictive or diagnostic inference from P\textsubscript{i} to P\textsubscript{j} (P\textsubscript{i} \Rightarrow P\textsubscript{j}) follows as ex-
plained in sect. 6.3. If there are n prototypical properties i.w.s. and m < n prototypical properties i.n.s., then the category C figures as mediator in a network of m⋅(n−1) direct predictive or diagnostic inferences P_i \Rightarrow P_j.

The cue validity of a property P for a category C is measured as the statistical probability of P among C-members divided through the statistical probability of P among members of the super-category C^+ in the underlying classification tree (cf. Rosch and Mervis 1975; Kleiber 1998: 52). Thus, the properties with high cue validity for category C are precisely the prototypical properties of C i.n.s. The cue validity of a category is defined as the sum of the cue validities of its prototypical properties i.n.s.. It follows that the cue validity of a category C increases with the number of its prototypical properties i.n.s. (cf. Rosch et al. 1976; Kleiber 1998: 63f). It was observed by Rosch et al. (1976) that the categories with highest cue validity are usually categories at a medium level of generality. These categories are called base categories – here are some examples (cf. Kleiber 1998: 62):

Superordinate categories: animal, fruit, furniture

Base categories: dog, apple, chair

Subordinate categories: poodle, golden delicious, folding chair

That categories with high cue validity such as 'dog' have medium level generality has the following evolution-theoretic explanation: their branch of ancestors in the tree of evolutionary descendence has a long, homogenous and category-specific selection history which produced many prototypical properties i.n.s. The selection history of more specific categories such as 'poodles' was too short to produce many prototypical properties i.n.s. of 'poodles'. In contrast, the selection history of 'animals' was very long but much too heterogeneous to produce many prototypical properties i.n.s. of 'animals'. Likewise for fruit and furniture.
In the remainder of this section I want to show how the objections 1-4 of sect. 4 can be
(dis)solved by an evolution-theoretic understanding of prototypes. The major objection 5, the
problem of non-compositionality, is treated in the final sections 6-9.

5.1 Analyticity and intersubjective meaning stability. The problem of intersubjective mean-
ing stability needs not to be solved by us; it is solved 'by nature'. If the environment has prototypi-
cality structure, then there is no need of defining kinds by an exhaustive list of properties. Indicating
some prototypical properties (e.g. 'have wings') is sufficient to grasp the underlying kind ('bird') and
discriminate it from other kinds ('mammal', 'fish') by the cognitive operations of induction and ab-
straction. Even if the process of concept acquisition via prototypes is different from person to per-
son it will nevertheless lead to intersubjectively stable meanings without any analytic definitions or
conventions – at least in most cases. This is only true because nature has shaped the property-
distribution of evolutionary systems in the way of sharp peaks as described below theses P2.1-2. As
a thought experiment, imagine that there would really exist a multitude of borderline cases between
different evolutionary species – say, pigeons with a tail and teeth and four legs, dogs with wings
and feathers, fish with wings and tails, etc. In such a possible world the process of concept learning
by prototype semantics would be hopelessly inefficient.

5.2 Typicality and Membership – Reasons for the Dual Theory of Meaning. There will always
be exceptional instances whose membership to a category is not clearly decided because they lack
some (or even many) of the prototypical properties of the category. In order to handle these more-
or-less atypical cases, or transition cases between different categories, one has to single out an ana-
lytic core meaning of the concept. For example, in order to decide whether swimming viviparous
animals (whales etc.) are mammals or fish, we need to fix an analytic core meaning of 'fish' and
'mammal'.

Singling out an analytic core meaning is certainly not the task of prototypes. For the functioning
of prototypes it is sufficient that exception cases are statistically rare – and this is granted by their
evolutionary basis. Fixing an analytic core meaning is partially achieved by common sense conventions, in order to stabilize communication. Only in rare cases, the analytic core meaning can be fixed by explicit definitions; much more often it is fixed, at least partially, by so-called *ostensive* 'definitions' ("this is a fish") 8, or by necessary semantic conditions ("mammals can't be fish"). In a more advanced way, fixing the meaning and deciding semantic borderline cases is the task of scientific *theories*. For example, the classification of species in modern biology is no longer expressed in terms of their prototypical properties, but in terms of their phylogenetic genealogy. This theoretical classification system is more coherent and unified; it agrees in many cases with the prototype classifications of common sense, but there exist also several differences which have caused many controversies in biology (cf. Ridley 1993: 369f).

The above argument is a clear reason for advocating the *dual* theory of meaning explained in sect. 4. Adopting this dual theory leads us to the next objection.

5.3 The Cognitive and Semantic Function of Prototypes. The cognitive function of prototypes has already been worked out: they enable fast and efficient predictive and diagnostic reasoning about evolutionary systems. The remaining question to be discussed is whether and why this cognitive function of prototypes should be understood as a *semantic* function, i.e. as being part of the *meaning* of a concept. I agree with Quine (1951) that in natural languages the separation between an analytic core meaning and a synthetic (world-dependent) knowledge about the concept's extension is neither sharp nor clear. Of course, one may *insist* that the meaning of a concept should be totally independent of world-knowledge, but such a decision would deprive the resulting notion of 'meaning' much of its psychological content. If one wants a psychologically realistic notion of meaning, which reflects that content which natural language speakers immediately associate when parsing the utterances of linguistic expressions (cf. Springer and Murphy 1991), then prototypes *should* be re-

---

8 This is closely related to Fodor's suggestion (1998, 95) that semantic "Fido"-Fido-principles ("dog" denotes dogs, etc.) are analytic.
arded as part of the meaning. Prototypes are also important for two further tasks which according to Fodor (1999: 176) a semantic theory should fulfill: they should be preserved by good translations, and they should be contained in a notion of content which is adequate for purposes of intentional explanation.

5.4 Concepts without Prototypes. Heterogeneous concepts of the sort explained in sect. 4.4 do not belong to the domain of evolutionary systems. Because of the restriction of the domain of prototype theory to evolutionary systems, these concepts are no longer a problem of prototype theory.

One may object that my evolution-theoretic account of prototypes now has the inverse problem: it is too restrictive. For example, not only kind concepts but also property concepts such as 'colour' have prototypes. However, in these cases there is usually a hidden evolutionary explanation in the background. For example, the similarity space of colours in human vision has its explanation in terms of the colour perception system of the human eye – which is, of course, an evolutionary system.

6. Formal versus Procedural Compositionality

We now turn to the problem of the compositionality of prototypes. Since not every complex concept has a prototype, we ask the question of compositionality in the following restricted form: provided a complex concept has a prototype at all, is the complex prototype (e.g. 'brown apple') compositionally determined by the prototypes of the constituent concepts ('brown' and 'apple') together with the syntactic operation (here, the adjective-noun-combination)? Before we can answer this question, we have to make an important distinction between formal and procedural compositionality.

Let E be an assumed set of grammatical expressions e∈E of a given language L which are built up from a set of atomic expressions A⊂E by a set Σ of syntactic rules of the form s: E^n→E. Let µ:E→M be a meaning function which assigns meanings m∈M to all grammatical expression of L,
and assigns a corresponding meaning function \( \mu_s \) to each syntactic rule \( s \). Then the principle of compositionality can be expressed by the following equation (cf. Hodges 2001):

\[
(\text{PC}): \mu(s(e_1, \ldots, e_n)) = \mu_s(\mu(e_1), \ldots, \mu(e_n)), \quad \text{for all } e_1, \ldots, e_n \in E \text{ and } s \in \Sigma.
\]

A given language \( L \) together with a given partial meaning function for \( L \)'s sentences \( \mu : \text{Sent}(L) \to M \) is called formally compositional if there exists a total compositional meaning function \( \mu \) for \( L \)'s expressions which preserves the meaning of \( L \)'s sentences. This idea of formal compositionality is based on Frege's context principle, according to which the meaning of a term (or atomic expression) of a language is determined by the term's contribution to the meaning of sentences containing it. It follows from the first extension theorem of Hodges (2001) that every language \( L \) whose sentence meanings are semantically functional, i.e. are preserved under substitution of synonymous subexpressions, is formally compositional.

Several authors have observed that formal compositionality is a rather weak condition (Zadrozny 1994), though it is certainly not empty (Westerstahl 1998). A counterexample to formal compositionality is

---

\( Proof: \) Hodges 1st extension theorem states that for every subclass of expressions \( E^* \subset E \) which is cofinal in \( E \) (i.e., every \( e \in E \) is subexpression of some \( e^* \in E^* \)) and which has a meaning function \( \mu^* : E^* \to M \) which is '1-compositional' and 'husserlian', there exists an extended meaning function \( \mu : E \to M \) which coincides with \( \mu^* \) over \( E^* \), is fully compositional and is unique up to isomorphism. We apply this theorem by letting \( E^* \) be the set of all sentences Sent(\( L \)). Then Hodges condition of '1-compositionality' coincides with the condition of preservation of sentence meanings under substitution of synonymous subexpressions of the same grammatical category (i.e., \( \mu(e) = \mu(e') \Rightarrow \mu(S) = \mu(S[e'/e]) \)). Hodges 2nd condition of 'husserlianity' requires that replacing synonymous subexpression preserves meaningfulness; this condition is met because 1-compositionality is assumed to hold for all sentences, i.e. whenever \( S \in \text{Sent}(L) \), then also \( S[e'/e] \in \text{Sent}(L) \), whence meaningfulness of \( S \) implies meaningfulness of \( S[e'/e] \) by 1-compositionality.
ositionality is only given when we encounter an *idiomatic expression* such as 'red herring': although 'herring' and 'an exemplar of Clupeidae' are semantically equivalent, 'a red exemplar of Clupeidae' means something different than 'a red herring'. In contrast, the typical problem of compositionality with prototypes, as exemplified in the *pet fish problem*, has nothing to do with formal compositionality. As long as 'pet fish' has a unique prototype, call it *protopetfish*, we can make *μ* formally compositional for 'pet fish' simply by setting *μ*('pet fish') = *μ*AN(*μ*('pet'),*μ*('fish')) := *protopetfish* (where *μ*AN is the meaning function underlying adjective-noun-combinations). However, the so-defined meaning function need not at all be computable.

To understand the *pet fish problem*, we need the stronger notion of procedural compositionality: a formally compositional meaning function *μ* is called *procedurally compositional* if *μ* is computable (i.e. recursive) – or even stronger, computable in a reasonable (i.e. polynomial) time. It is the violation of this condition which was the target of the criticism of prototype's non-compositionality. For example, the prototypical pet fish cannot be computed in an obvious way from the prototype of a fish and that of a pet.

### 7. Refined Accounts to Prototype Composition and their Limitations

In many cases of adjective-noun combinations, the adjective does not correspond to a conjunctive term but functions as a *modifier* of the noun. For example, that *x* is a skillful *F* does not mean that *x* is an *F* and *x* is skillful, but rather that *x* is an *F* with special *F*-skills. Several authors have at-

---

10 For a similar suggestion cf. Kracht (2001). A second condition for procedural compositionality was introduced in Schurz (2005b, 282). This condition requires that the given algorithm computes *μ* in a *bottom-up* fashion. This is needed for handling the semantics of theoretical terms, but not for prototype semantics. Procedural compositionality becomes equivalent with *epistemic compositionality* in the sense of Bonnay (2005, 43) if the algorithm A is used by the epistemic subject.
tempted to construct compositional meaning functions which model the meaning of adjective-noun combinations by certain modifications of the noun meaning. An advanced account of this sort is the selective modification model of Smith et al. (1988).\footnote{Another account is the adjective-noun calibration theory of Kamp and Partee (1995, 164f), which has been criticized by Osherson and Smith (1997, 197f), and the adjective-noun combination model of Wisniewski (1997). Hampton (1987) has developed a concept integration theory which includes the influence of world-knowledge.} Recall the prototype frame of the noun 'apple' as explained in sect. 4. If the noun 'apple' is combined with an adjective such as 'red', then the prototype of the combined concept 'red apple' is obtained by the following modification operation on the prototype frame of 'apple': (a) the adjective 'red' shifts all votes in the corresponding attribute 'colour' to its own value, (b) it increases the diagnosticity of the colour attribute, and finally (c) it leaves the structure of all the other attributes unchanged:

\begin{table}[h!]
\centering
\begin{tabular}{|l|l|l|l|}
\hline
attribute (diagnosticity) & values (votes) \\
\hline
colour (2) & red (30) & green (0) & brown (0) \\
shape (0.5) & round (15) & square (0) & cylindrical (5) \\
texture (0.25) & smooth (25) & rough (5) & bumpy (0) \\
\hline
\end{tabular}
\end{table}

It has been experimentally confirmed by Smith et al. (1988: 364ff) that this model works in many cases, but not in all. The major restriction of any selective modification model of this sort is the following: the modifying adjective must not express an exceptional property. Otherwise the rule of default-inheritance of the remainder prototypical properties which is assumed in condition (c) of the selective retention model may get violated. For example, it is a prototypical property of apples to have smooth texture, and this prototypical property is inherited by normal (red or yellow) apples,
but brown apples\textsuperscript{12} have rough texture; or it is a prototypical property of apples to taste sweet, and this prototypical property is inherited by red or yellow apples, but green apples are usually not ripe and hence taste sour. To give another example, birds usually can fly, and this prototypical property is inherited by all normal birds, but birds in the Antarctic do not fly because they are penguins.

In conclusion, whenever the modifying adjective expresses an exceptional property, then the rule of default-inheritance may be violated, because the exceptional property causes shifts in the frequency-distributions over the values of the other attributes, in a way which is not solely determined by the meaning of the noun and the adjective but depends on specific pieces world-knowledge, such that brown apples have a rough surface, green apples are usually unripe, and almost all birds in the Antarctic are penguins who cannot fly. It follows that the prototypes of complex nouns whose adjectives express exceptional properties are not (fully) compositional in the procedural sense (although they are formally compositional as long as compound prototypes are functionally unique). This fact is also acknowledged by Smith et al. (1988: 386f) and further supported by Hampton and Jönsson (this volume). A further problem of the selective modification model arises when the modifying adjective is 'new' in the sense that it does not belong to any of the value spaces of the prototypical attributes in the noun frame. Smith et al. (1988: 386ff) suggest that in cases in which the new adjective is 'neutral', an appropriate attribute may be added to the prototype frame of the noun. In cases in which the new adjective expresses an exceptional property, it will again change the value frequencies of the other attributes (e.g. 'killer dog'). The exceptional adjective may even necessitate the inclusion of further attribute-values in the frame of the compound noun phrase (e.g. 'victims of a killer dog').

8. Explaining the Composition of Prototypes by Probabilistic Default Reasoning

\textsuperscript{12} Contrary to Smith et al. (1988, 360f), 'brown apples' exist; their German name is 'leather apples'.
All of the discussed logical properties of prototypes can be explained by the rules of *probabilistic default reasoning*, in short: PDR. These rules, which are summarized in the system $\mathbf{P}$, go back to Adams (1975) and have been further developed by various researchers on non-monotonic logic.

Instead of presenting system $\mathbf{P}$ in all of its variants (for overviews cf. Schurz 1998, 2005a) I explain only those aspects which are important for prototypes. Three general features of PDR are the following:

1.) PDR interprets normic conditions $'A \Rightarrow B'$ as *high conditional probabilities* $'P(B|A) = \text{high}'$ (recall sect. 1). In the evolution-theoretic understanding 'P' is regarded as a *statistical* (rather than a merely subjective) probability (recall sect. 2). The so-called *uncertainty* of $A \Rightarrow B$, $U(B|A)$, is defined as $1 - P(B|A)$.

2.) The rules of PDR infer normic (conclusion) conditionals from given (premise) conditionals; thus these rules have the format $'A_1 \Rightarrow B_1, A_2 \Rightarrow B_2, \ldots / A_n \Rightarrow B_n'$ (the stroke '/' separates the premises from the conclusion). The PDR-rules for normic conditionals ($\Rightarrow$) are *weaker* than the classical rules for strict (exceptionless) conditionals ($\rightarrow$). For example, normic conditionals satisfy merely the rule of cautious transitivity $'A \Rightarrow B, A&B \Rightarrow C / A \Rightarrow C'$, while strict conditionals satisfy full transitivity $'A \rightarrow B, B \rightarrow C / A \rightarrow C'$.

3.) The semantic criterion of *correctness* which underlies the rules of PDR is the *preservation of high conditional probability*. Although high conditional probability is qualitatively preserved by PDR, a certain probability loss is unavoidable. This probability loss increases with the *number* of premises, but it is *controlled* by the following (mathematically proved) condition: a (conclusion) conditional can be inferred in calculus $\mathbf{P}$ from given (premise) conditionals if and only if in all probability models the uncertainty of the conclusion conditional is less-or-equal than the sum of the uncertainties of the premise conditionals. E.g., for cautious transitivity this means that if $P(B|A) \geq 1-\varepsilon_1$, and $P(C|A&B) \geq 1-\varepsilon_2$, then $P(C|A) \geq 1-\varepsilon_1-\varepsilon_2$. Similar probabilistic inequalities hold for *all* inferences which can be drawn within system $\mathbf{P}$; in other words, the system $\mathbf{P}$ is under full probabil-
istic control.

Three central aspects of reasoning with prototypes can be explained by PDR as follows.

8.1 Predictive and diagnostic reasoning with prototypes. A prototypical property i.w.s. \( P_i \) of an underlying category C supports a normic conditionals of the form (i) \( C \Rightarrow P_i \). A prototypical property i.n.s. \( P_k \) supports in addition to (ii) \( C \Rightarrow P_k \) also the inverse conditional (iii) \( P_k \Rightarrow C \) – i.e., the prototypical property i.n.s \( P_k \) is indicative for the corresponding category. From (i), (ii) and (iii) the direct predictive inference (iv) \( P_k \Rightarrow P_i \) can be inferred by PDR in the following subtle way. Neither full transitivity (\( P_k \Rightarrow C, C \Rightarrow P_i \) → \( P_k \Rightarrow P_i \)) nor full monotonicity (\( C \Rightarrow P_i / C \& P_k \Rightarrow P_i \)) are rules of the system \( P \). However, from (i) and (ii), (v): \( C \& P_k \Rightarrow P_i \) follows by the \( P \)-rule of cautious monotonicity; and from (iii) and (v), the intended conclusion (iv) follows by the above-mentioned \( P \)-rule of cautious transitivity. In other words, predictive and diagnostic reasoning based on prototypical properties is only reliable as long as those properties which are used as indicators for the corresponding category (iii) are at the same time prototypical i.w.s., i.e. are statistically frequent (ii).

8.2 Default inheritance of prototypes – a correction of Connolly et al. (2007). The rule of default inheritance of prototypical properties, in short (DP), says formally that \( C \Rightarrow P \) implies \( CA \Rightarrow P \), where CA is a subkind of the category C which obtained by a modifying adjective A. Logically speaking, the subkind CA can always be equivalently represented as a conjunction 'C&X' for some X. Therefore the rule (DP) is nothing but an unrestrained monotonicity rule (\( C \Rightarrow P / C \& A \Rightarrow P \)). It is clear that the rule (DP) cannot be generally probabilistically safe, because of the well-known fact that conditional probabilities are non-monotonic, i.e. a high value of \( P(B|A) \) does not entail a high value of \( P(B|A \& C) \). However, this does not mean that this rule is generally unreliable, as Connolly et al. (2007) have argued. It is true that the prior probability that an arbitrary object x is a CA is smaller-or-equal than that an arbitrary object x is a C. But this does of course not imply, that "the probability that an arbitrary AN is a good instance of N is smaller-or-equal than the probability that an arbitrary N is a good instance of N", which is the claim of Connolly et al. (2007: 8). That an arbi-
trary AN is a good instance of N means that 'if something is an AN; then is a GN', where 'GN' stands for 'good N' and means the possession of many prototypical properties of Ns. If the conditional (if X, then Y) is interpreted as a material implication ($\neg X \lor Y$), then the probabilistic relation would even be the inverse of what is claimed by Connolly et al. – then $P(\text{if AN, then GN}) \geq P(\text{if N, then GN})$ would hold. However, the conditional assertion has to be interpreted not as a material conditional but as a high conditional probability, and in this case, there exists not general world-independent relation between the value of $P(\text{GN|N})$ and $P(\text{GN|AN})$: depending on the facts of the world, $P(\text{GN|AN})$ can be greater, smaller or equal to $P(\text{GN|N})$.

8.3 Default inheritance and the rule of rational monotonicity. There exists a PDR-rule of the extended system $\mathbf{P}^+$ (cf. Schurz 1998) which expresses precisely what we have worked out before, namely that default-inheritance by subkinds is reliable if these subkinds are non-exceptional in the statistical sense. This is the rule of rational monotonicity which says the following:

*Rational monotonicity (RM):* From $C \Rightarrow P$ and $\neg(C \Rightarrow \neg A)$ infer $C \& A \Rightarrow P$.

In words: if P is a prototypical property of C, and the subcategory CA is not exceptional (i.e. Cs are not normally not CAs), then P is probabilistically inherited to the subcategory CA.

(RM) is stronger than the rule of 'cautious monotonicity' "CM: from $C \Rightarrow P$ and $C \Rightarrow A$ infer $C \& A \Rightarrow P"$, because the second premise of (RM) is much weaker than the second premise of (CM). The rule (RM) satisfies the probabilistic inequality $U(P|C\&A) \leq U(P|C) / P(A|C)$; i.e. if $P(A|C)$ is not small then $U(P|C\&A)$ will not be much greater than $U(P|C)$.

---

13 Connolly at al. (2007) call the rule (DP) 'DS' for 'default to stereotypes'. The inequality $P(\text{GN|AN}) \leq P(\text{GN|N})$ holds for the special probability distribution by which Connolly et al. illustrate their claim (2007, 9); but it does not hold in general, as they assume.

14 $\neg(C \Rightarrow \neg A)$ requires that $P(A|C)$ is not small, while $C \Rightarrow A$ requires that $P(A|C)$ is high.
Rule (RM) tells us that when the subclass-forming property is not a statistically rare, the rule (DP) is reliable. But recall from sect. 8.2 that the inverse direction does not generally hold: not all statistically rare properties cause exceptions, i.e. induce changes in other prototypical properties – those who don't are called 'neutral'. For example "living in Duesseldorf" is a statistically rare but neutral property of Europeans. Nevertheless, it hard to see how a semantic computation mechanism could discern, without any empirical information, those cases of statistically rare subclasses which cause exceptions from those which are neutral. Therefore I conclude that general composition procedures are reliable only when the modifying adjectives to which they are applied are not statistically rare.

The rule (RM) explains also the experimental results of Connolly et al. (2007: 10ff.). Very briefly, Connolly et al. (2007) found out that the test persons' estimations of the likelihood with which members of a given kind have a prototypical property (e.g. ducks have webbed feet) decreases if the kind is constrained by more and more neutral adjectives (e.g. baby Peruvian ducks have webbed feet). Assuming the evolution-theoretic hypothesis of sect. 2, that humans are quite fit in reasoning with normic conditionals, the probabilistic explanation of this result is as follows: the more adjectives $A_1A_2\ldots$ modify the noun $C$, the smaller the conditional probability $P(A_1A_2\ldots|C)$ gets, whence rule (PD) becomes more and more unreliable. There is nevertheless a correlation between a property $P$ being prototypical for a given category $C$ and its tendency to be inherited to restricted subcategories $CA$, but this correlation sinks within increasing specificity of the CA-subcategory (cf. Hampton and Jönsson, this volume).

9. Conclusion: Prototypes are Semi-Compositional

The forgoing considerations support the conjecture that prototypes are semi-compositional in the following sense: there exist unboundedly (i.e., potentially infinitely) many combinations of nouns
with non-exceptional adjectives which satisfy the rule (DP) and hence are compositional, but presumably there also exist unboundedly many combinations of nouns with exceptional adjectives which violate (DP) and hence are non-compositional. Although I don't have a proof that there exist unboundedly many exceptional adjectives I can think of no principled reason why their number should be finitely bounded, because exceptional properties can be conjoined with other (non-exceptional or exceptional) properties to form new exceptional properties, iteratively up to unbounded complexity. What does this imply for the explanation of the facts of productivity and systematicity of meaning in natural language, for which compositionality was claimed to be the best if not the only explanation?

An insightful analysis of the connection between productivity (systematicity) and compositionality has been suggested by Robbinns (2002). He argues that for the explanation of productivity one need not assume that conceptual meanings (contents) compose always – it is enough that they compose in in(de)finitely many cases (ibid.: 317, 321f). According to the hypothesis of semi-compositionality, this is indeed the case. So we can conclude that semi-compositionality yields an equally good explanation of productivity and hence is not refuted by Fodor's arguments. On the other hand, since semi-compositionality entails the existence of unboundedly many non-compositional cases, non-compositionality cannot be dealt away with a finite list of idiomatic exceptions, but is a genuine feature of prototype semantics in natural language.

REFERENCES


McCarthy, J. (1986). 'Application of Circumscription to Formalizing Common-Sense Knowledge',


