Foreword

Mariano L. Bianca e Paolo Piccari

> How comes [the mind] to be furnished? Whence comes it by that vast store which the busy and boundless fancy of man painted on it with an almost endless variety? Whence as it all the materials of reason and knowledge? To this I answer, in one word, from experience.
>
> J. Locke, *An Essay Concerning Human Understanding* (1690)

In recent years the debate on concepts has been enforced as advances in cognitive science have provided philosophers with new tools refining the traditional dispute. As we know, different hypotheses have been suggested, whose main goal is to explain the nature of concept and its structure. Philosophers have got advantage from empirical studies in such different fields as developmental psychology, evolutionary psychology, cognitive anthropology, neurosciences, linguistics, and ethology. This Special Issue of *Anthropology and Philosophy* devoted to the *Concept Formation* attempts to offer some contributions to the contemporary research and discussion from different perspectives.

In *Conceptual Framework: A Frequency Model* Mariano L. Bianca and Paolo Piccari focus their attention on the nature of empirical concepts. Such concepts are considered as a *conceptual framework* represented by a five–dimensional vector in which, in addition to the merely perceptive content (the identitive and specifying perceptive attributes), non perceptive contents (semantic reference and various significances assigned to concepts by single individuals) are analysed. This model does not consider empirical concepts as the simple result of a generalization conducted on the basis of different perceptive instances, but highlights the relevance of non perceptive contents in their formation.

In *Conjunctions of Social Categories Considered from Different Points of View* James A. Hampton, Margaret Dillane, Laura Oren and Louise Worgan argue that conjunctions of divergent social categories may elicit emergent attributes to render the composite concept more coherent. On the one hand, social categories can be combined in an integrative fashion, taking the positive and negative attributes of each category and combining them into a novel composite prototype. In order to identify a set intersection of instances in the world, it is necessary to create a set union of the criteria that identify them. On the other hand, they have discovered that in certain circumstances, people will resist this integration. Throughout two experiments they showed that: a) when adopting the point of view of one constituent category, people tended to combine the concepts antagonistically, meaning that they attributed to members of the conjunction the more negative aspects of the opposing category; b) this polarizing effect was reduced when the point of view category was itself unusual.

In *The Hidden Strengths of Weak Theories* Frank Keil stresses there has been a strong tradition of assuming concepts as embedded within larger systems of beliefs that help to articulate their structure. He argues there is a newly emerging concern that is much more challenging to address — people's intuitive theories seem to be so impoverished it is difficult to see how they could provide the necessary structure to explain differences between concepts and how they might form in development. One response to this recent challenge is to abandon all views of concept structure as being related to people's intuitive theories and see concepts as essentially structure–free atoms. The alternative proposed in this paper argues that our very weak theories might in fact do a great deal of work in explaining how we form concepts and are able to use them to successfully refer.

In *Concepts as General Representations in Situated Theories* Elisabetta Lalumera addresses the issue of the role of concepts as representations of general knowledge, which seems intuitive, but it is seldom explained. She identifies two kinds of general knowledge, namely, constitutively general (possessed by all members of a category) and behaviourally general (that can be applied to all members of a category). First Lalumera reviews the ways in which traditional theories of concepts have coped with generality, and then focuses on situated or 'embodied' theories, which present themselves as highly revisionary with respect to other models.

In *Against Hybrid Theories of Concepts* Edouard Machery and Selja Seppälä argue that the psychologists of concepts' traditional assumption that there are many properties common to all concepts has been subject to devastating critiques in psychology and in the philosophy of psychology. In this article, they compare two competing approaches, the Heterogeneity Hypothesis and the hybrid theories of concepts, and they present an empirical argument that tentatively supports the former over the latter.

From our point of view these papers can really contribute to develop further studies on the matter. We wish to thank the Authors of the papers who accepted our proposal so allowing the realization of the Special Issue on *Concept Formation*.

MARIANO L. BIANCA
PAOLO PICCARI

Conceptual Framework: a Frequency Model

Mariano L. Bianca – Paolo Piccari

Abstract In this paper we focus our attention on the nature of empirical concepts. These concepts are considered as a conceptual framework represented by a five–dimensional vector in which, in addition to the merely perceptive content (the identitive and specifying perceptive attributes), non perceptive contents (semantic reference and various significances assigned to concepts by single individuals) are analyzed. This model does not consider empirical concepts as the simple result of a generalization conducted on the basis of different perceptive instances, but highlights the relevance of non perceptive contents in their formation.

Key words Concept, generalization, prototype, mental representation, conceptual framework, visual perception.

1. Overview

Over the last 40 years the study of concepts, particularly within semiotic and cognitive disciplines, has become one of the most engaging fields of research producing a wealth of relevant findings, despite the fact that none of the descriptive or explicative models of conceptual structures has been proved adequate. The question posed is the same one that Locke and Husserl had already formulated: What are the mental processes take place when we speak, for example, about cats? The underlying question then being: What is a cat? This question has been overlooked in the analytical tradition, according to which it is not just important to ask oneself what a cat is but rather to verify whether the proposition 'a cat is an animal' is true or false. Furthermore: Why do individuals break up daily experience into discrete units by assigning them names? Why, then, 'classify' or 'categorize'? Why group objects or events into classes, or, to use a term which is widely diffused in the cognitive sciences, into 'categories'? Why, then, formulate concepts?

Frege, contrary to every form of psychologism, asserted that a concept is objective, that is, it does not depend on our mental contents, and thus a proposition such as "the number 3 is a prime number" exists independently of the fact that we wake, sleep, live or not: something that holds true always will be and is irrelevant to the existence, present or future, of beings which recognize or do not recognize that truth (Frege 1891a). Furthermore, Frege, considering concept as function and not as a property, brought about a turning point of great relevance in philosophy, extending the application of the concept of function beyond a strictly mathematical realm. Arguments and values of a function are not just numbers, but also any kind of "objects" (the term "object" has been used by Frege). Therefore, in formulas such as f(x) = y not only can numeric expression occupy the place of arguments and values, but any sign or symbol. One example is the following: "x is the inventor of penicillin"; whatever object takes place of x, the value of the function is true or false.

According to Frege, therefore, "a concept is a function whose value is always a truth–value" (Frege 1891b). He went on to clarify that a concept is "predicative," that is, it is the meaning of a grammatical predicate (Frege 1892). The concept *dog*, for example, is designated by the expression 'x is a dog'; in this case the objects which can constitute the argument are designated by expressions collocated in subject positions and in this case x can be substituted by terms like 'bull dog', fox terrier', etc.

Thus Frege moved the first step towards resolving the notion of concept into that of meaning, which constituted the logical paradigm of reference for the most part of contemporary philosophical reflection on the nature of concepts up to the middle of the last century until Wittgenstein's first observations were formulated (Wittgenstein 1953). According to Wittgenstein, daily concepts have a weaker structure than the *definitional* ones; later on, Eleanor Rosch's studies in cognitive psychology definitively demonstrated the unsustainability of the reduction of the notion of concept to that of meaning (Rosch 1975a, 1975b, 1977, 1978).

From a cognitive point of view two functions of concept have been considered. Firstly, that of favoring cognitive economy, because it is through the codification of experience that concepts allow for the diminution of the quantity of information that must be remembered. For example, instead of remembering every exemplar of *cat* one has ever encountered, we only remember one or at most some of those representing the class of cats. We may even have a more abstract representation, independent of the memory of each experienced exemplar.

The other important function of concepts is that of favoring inference: after having classified a cat as such, many of its specific attributes can be inferred — among which, for example, that of meowing — although these are not immediately objects of perception (Girotto–Legrenzi 1999:91).

The term 'concept', in general, has a wide meaning and can refer to any object, abstract or concrete, near or far, universal or individual. It is possible to have a concept of a fork, of the number 7, of humans and of God, of genus and species (the so–called *universals*) as a specific reality, such as a historic or artistic period like the French Revolution and the Baroque.

Although it is generally indicated by a name, a concept is not the name insofar as it is possible that different names can indicate the same concept or different concepts can erroneously be indicated by the same name.

In light of the most recent philosophical and psychological literature pointing towards new and interesting research perspectives, in this article we intend to examine in particular the formation of empirical concepts: those concepts which refer to a certain number of perceptive experiences related to objects or events considered as belonging to the same class. To reach this goal, we will briefly analyze the classical theory of concepts and the theory of prototypes, examining some of the more relevant questions, for example, how does an individual formulate the concept of 'car' on the basis of an empirical generalization after having inducted the attributes common to each car s/he has encountered?

Following this outline, empirical concepts are considered as *empirical generalizations* derived from information contained in the mental representations corresponding to a certain number of object/events which have been perceived and can be placed in a class because they share some attributes.

This paper will not take into consideration other types of concepts like *theoretical* ones, which are not derived, at least not directly, from empirical generalizations such as scientific concepts like the Big Bang, black energy, relativity, or philosophical concepts dealing with substance, transcendence, or theological concepts about God or the soul. We will also not analyze *lexical* concepts for they do not derive from empirical generalization but rather from a linguistic description as one would encounter when learning the meaning of a lexeme found in the dictionary.

In section 2 we will examine the classical theory of concepts and the theory of prototypes; in section 3 we will consider the processes of generalization and the formation of empirical concepts; in section 4 we will analyze the structure of these concepts; in section 5 we will take into consideration visual concepts and, finally, in section 6 we will formulate some conclusions.

2. Classical Theory of Concepts and Theory of Prototypes

The *classical theory* of concepts from antiquity through the 1870s was universally accepted in the philosophical studies. Although it has been supplanted by other more recent theories, due in part to some simplistic and superficial interpretations, it is not possible to seriously deal with the topic of the formation and of the function of concepts without referring to classical theory, according to which concepts are complex mental representations whose information allow us to specify the conditions necessary and sufficient for their application. Consider, for example, the concept 'bachelor': this is a complex mental representation whose identifying attributes are 'unmarried, 'adult' and 'male'. An object is part of the *extension of* 'bachelor' only in the case that it satisfies these attributes; in other words, in order for a man to be classified as a 'bachelor' he must have all of the identifying attributes (*criterion of necessity*), which are also sufficient for that classification (*criterion of sufficiency*). Hence, in the case of the concept 'bachelor', the intension is 'to be an unmarried adult male', while the extension is represented by all of the exemplars (the members of the class) who find themselves in this condition.

What happens when one classifies objects according to the classical theory? To classify an object like 'dog', we must break up the concept of 'dog' and verify whether or not its attributes can be applied to the object in question. If all the attributes of that object are present in the concept, then the object can be considered a dog; if only one of those attributes is not applicable, then the object cannot be considered a dog.

The aporias and counterexamples that stem from the classical theory are so well known that we shall forego reviewing them and refer the reader directly to the vast literature dedicated to this topic (Margolis–

Laurence 1999; Murphy 2002; Prinz 2002; Machery 2009). In any case, as an example we might refer back to the concept 'bachelor': in its rigid definition, is it applicable to the Pope — a human being, of the male gender, unmarried? He cannot be defined as a bachelor because he cannot enter into marriage for priests of Latin rites are sworn to celibacy.

In this case the application of the concept is bound by a cognitive model regarding the typical attributes of 'bachelor' and its consequent behaviors or those enacted about him: a bachelor, for example, is an individual who can be presented to a female friend and might become her partner (Lakoff 1987). Furthermore, it is also important to underline that amongst members of the same class there could be a diversity in reference to their degree of *typicality*, that is, the degree to which each object represents a determinate class; for example, a finch would be classified as a bird more rapidly that would a penguin. On the contrary, at the base of the classical theory, if an object satisfies the conditions of affiliation with a class, it does so independently of the degree of typicality because members of a class possess an equal degree of pertinence to the same class.

In spite of these difficulties, it must be stressed that the classical theory offers an adequate explanation of the concept learning process for it is true that we learn a concept only after having acquired the attributes that make it up. In addition, that theory satisfies the criterion of economy for a single representation is used in place of an entire class.

Unlike the classical theory, the theory of prototypes holds that a class of objects is not defined by a set of necessary and sufficient conditions, but by the *best exemplar*, that is, by the *prototype*. If indeed there were a set of attributes upon which one could precisely define a class, and if those attributes pertained to all the members of the class, then each member should be equally a good exemplar of the class without differences regarding its degree of typicality.

Upon closer examination, the descriptive model of conceptual thought whose formation of concepts is meant only as a process of generalization of experiences and representations of objects in the world, is incomplete and inadequate. For example, if we look at and experience different tables, we could formulate, through inductive reasoning, a *typical exemplar* of the concept 'table' consisting of the attributes 'being a piece of furniture', 'having a plane parallel to the ground', 'having one or more legs'.

Moreover, a typical exemplar, a single *prototype*, is not capable of supplying adequate information about the diversity in some attributes amongst the members of a class of objects it refers to. For example, does an 'ideal bird' exist as representative of all birds, large and small, black and green, winged and wingless, carnivore and herbivore? It seems unlikely that a single generalization could encompass all of these different attributes.

It is important to note that from an epistemological point of view, the empirical generalizations involved in the concept formation are mostly referred to objects which have the same attributes and as such can be grouped in the same class: for example, the class of boats, the class of stars, the class of birds and so forth. Hence, empirical generalizations are expressed in propositions which refer to a class of objects of which some are perceived on different experiences.

Through empirical generalization, formulated by an inductive process, it is possible to describe the common attributes of a group of objects thereby establishing a class. More precisely, empirical generalization, for gnoseological purpose, allow to group a large number of 'cases' on the basis of common attributes which can be assigned to each one of them. Thus, the processes of empirical generalization are fundamental for acting in the world and cognitively approaching it, classifying objects or phenomena, whether experienced or yet to be experienced, whose common attributes allow them to be considered as belonging to a class of objects.

3. Generalization Process

3.1. Mental Representations

How is the process of generalization which leads to the formation of empirical concepts carried out? In order to answer this question we must introduce the notion of *mental representation* and explain briefly its nature and formation. Mental representation is one of the functional modalities of knowledge of the world by way of the use of sense organs or rather *sensorial transducers*.

The term 'representation' refers to the fact that sense data are elaborated and capable of generating one or more mental configurations whose content (information) is, though not always entirely, *isomorphic* to the structure of the world it represents. The mental representation, relative to every type of perception (visual, auditory, etc.), is a neuromental structure that reports the information coming from an object or event in the world (see Figure 1).

It has a double function: firstly, it is *cognitive*, that is, it generates *reliable* and *adequate* knowledge of the world; secondly, it is *operative*, that is, it triggers various forms of reaction or behavior, useful to acting and operating in the world, and a series of mental configurations such as thoughts, emotions etc. (Bianca 2009: 481–482).

Let us consider the case of visual perception. Take, for example, a book on a table. What happens when we perceive it? The photoreceptors of our retina (cones and rods) perceive visual stimuli (photons) emanating from the book (stimulus source). At this point, the codification of the information coming from the stimulus source takes place and the following neurophysiologic process (*structural transposition)* give rise to the mental representation, in this case the mental image of the book. A structural transposition is a complex neurophysiologic process which allows the transcription of reception of the stimulus by way of specific neurochemical codes belonging to the structure and the functioning of the central and peripheral nervous systems and generates a structure (the representation) which stands in mind for the structure of the object emitting the stimulus. The mental representation carries in a codified manner, though not always entirely, the attributes of the structure of the source of the stimulus (Bianca 2005: 90–95).

Mental images, for example, originating from visual experience, are sufficiently reliable representations of physical objects and as such are those which are prevalently utilized to epistemically access to the world and act within it (Bianca 2009: 27). Thus, a mental representation, which

Perceptive stimulus
↓
Flow of the information from the stimulus to the sensorial organs
↓
Change in the state of the sensorial transducers and neurochemical codification
↓
Elaboration in different cortical and non–cortical areas of information coming from sensorial transducers
↓
Mental Representation

Figure 1. Mental Representation Process.

is the result of neuronal processes brought about by sensory stimuli capable of triggering the activation of many neuronal networks, *isomorphically carries* the structure of the objects of the world, generates different mental configurations and can provoke specific behaviors and actions. For example, in the brain of a gazelle the representation of 'lion', which stands for a lion in the world, is the cerebral configuration that allows her to trigger a neurophysiologic configuration which corresponds to escape, and as such, to the act of escaping.

The previous description refers to a cognitive process capable of formulating a representation. This is the final step of an elaboration in which information transmitted by a stimulus coming from an external source is processed in different areas of the brain. As such, the representation provides knowledge that is *reliable* and *adequate* about the object (which *it stands for)* throughout the elaboration of information coming to the sense receptors from the stimulus, thereby triggering actions conforming to the state of the subject and the structure of the object. The aggressive stance of the lion, for example, is not mistaken by a gazelle as a gesture of submission, but is taken for what it really is; thus, the adequate action, that of escape, is triggered in the gazelle by aggressive stance of the lion. In this case, the cerebral representation of the lion in the brain of the gazelle is both *reliable* and *adequate*; reliable because it carries the attributes of the lion and *adequate* because it allows for a certain type of behavioral response, consonant with the received information, to be put into action.

As far as the formulating process of mental representations in humans is concerned, it is necessary to observe that a human mental representation, for example a visual perception (a visual image), is not just a reliable and adequate representation of one or more objects of the world, but bearing a collection of attributes is an 'empirical generalization' based on perceptive instances of similar objects which are so assembled in to a class. In which case, it can be considered as a conceptual structure (or concept) referring to an object or a class of objects; a concept that might be useful for classifying experienced objects in successive instances.

Furthermore, if one considers that perceptive experience of the human species is prevalently visual in nature, one should conclude that most of the concepts derived from empirical generalization processes are of a visual nature and in smaller number those derived by auditory, tactile, olfactory and gustative experiences.

3.2. Two kinds of generalization process

The formation of a mental representation is the process preceding that of generalization insofar as the empirically acquired information is elaborated in a representational form. Therefore, for example, it would be possible inductively form the concept 'book' on the basis of different visual perceptions (or representations) of the book which refer to the single perceptive visual instances of the 'book' (see Figures 2 and 3).

The inductive process is a generalization which refers to the perceptive instances of type X on the basis of the trial of a limited number of perceptive instances of type X. According to the definition of Aristotle in *Topics* (I, 12, 105 to 11), induction is a process in which details lead to universals, that is, the kind of reasoning that formulates a general statement referring to all cases which possess similar characters from a group of particular statements referring to single cases.

In inductive reasoning there are three distinguishable parts: a) the *base*, b) the *passage*, c) the *conclusion*. The first part contains the statements that refer to single actually experienced instances; the second, statements that are not formulated but are referred to possible other perceptive non–experienced instances of type X; finally, the third contains one or more statements in the form of generalization through which one states or assumes that the attributes evident for a few perceptive instances of type X experienced count for all of the analogous perceptive instances of type X not experienced, and therefore one concludes that they count for all of the possible instances of type X.

In the description of the process of generalization which gives rise to the formulation of empirical concepts we mean to make reference to induction in the statistical–probability sense expressed by the following rule: when we have observed a sample of objects α and found that the frequency of the objects β between them is f, we assume that $P(\alpha, \beta) = f$, that is, the probability that α is β is f, calculated on the basis of the statistical frequency of β in α.

In the case in which a specific attribute occurs in a wide *proportion* of experienced objects of type X, one can assume that attribute is common to all of the other objects of the same type though not experienced, except when proven otherwise. When the proportion is the same in 100% of the objects experienced, that is when the attribute occurs in all of them, there is a *uniform* or *complete* generalization. This is the case when we state that "all men are mortals" due to the fact that being mortal has always been found to be a constant in association with humankind. However, when the numeric value of that specific proportion is used as a measurement of possibility that the attribute in question will occurs in a new perceptive instance, there is a *judgment of probability*. As such, it is adequate to specify that uniform generalization or the judgment of probability are aspects of the statistical generalization based on the frequency of occurrences of an attribute in a sample of observed cases. The probability p(A) of the occurrence of an attribute identifying a class in a new perceptive instance like, for example, 'blackness' in a class of crows, is always so that $0 \leq p(A) \leq 1$; $p(A) = 0$ if and only if A is the impossible event, $(A) = 1$ if and only if an event is certain.

3.2.1. Generalization e pluribus and generalization ex uno

There are two forms of perceptive generalization, also in the statistical sense: *e pluribus* generalization and *ex uno* generalization (Bianca 2009: 261–265). *E pluribus* generalizations contain information that is attributive and referable, following the inductive process, to all of the objects of a given class (the experienced objects that we consider in the inductive process and the objects yet to be experienced). The *e pluribus* generalization can therefore be considered a concept that contains several attributes possessed by all (or nearly) of the objects of a class. It is preserved in the memory as a *perceptive type* and is useful for classifying the objects in successive perceptive experiences (Fig. 2).

However, in the case of *ex uno* generalizations certain recurrent attributes are assigned to a determinate object or event that has been experienced many times. This generalization identifies the single object and is accepted as *referable in every case* in which that single object or event has been or could be experienced or perceived. The *ex uno* generalization is a neuromental process that carries all of the attributes individuated in different perceptive instances of the same object, thus formulating a *perceptive type*, and allows that a new perceptive instance of that object be subsumed into that perceptive type (Fig 3).

Ex uno generalizations refer to a class of perceptive instances of the same object experienced in different instances, as for example: 'my university office yesterday evening,' 'my university office after the daily cleaning', 'my university office with a new computer', hence, instances which are experientially different but refer to the same object.

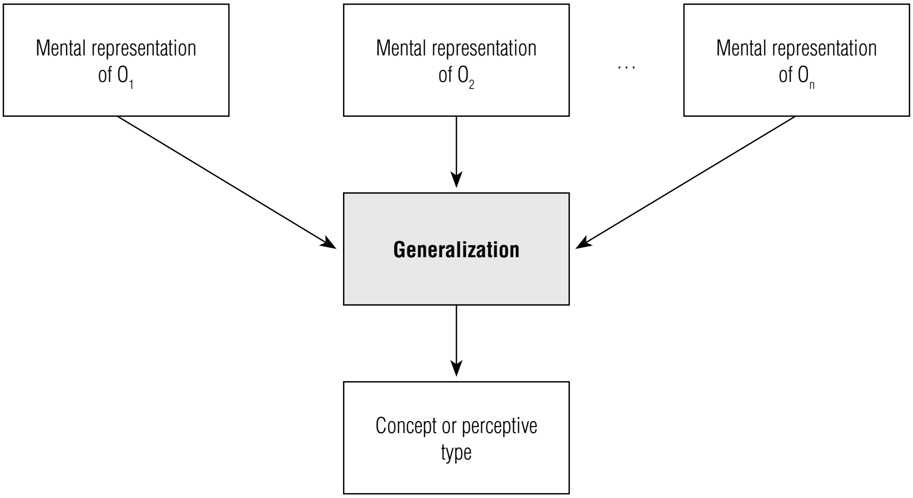

Figure 2. E pluribus generalization (O = Object).

Thus, we can define this generalization according to the following form: given a certain number of perceptive instances of an object/event a number of attributes gets assigned to it which are appliable to the perceptive instances of this object, although not all of them are appliable to every instance. In other words, the *ex uno* generalization identifies the single object

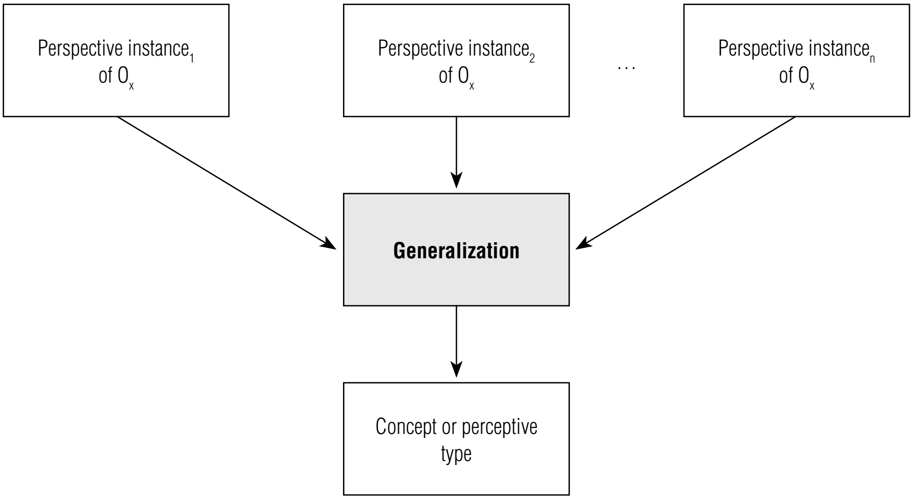

Figure 3. Ex uno Generalization (O = Object)

and makes reference to all of the perceptive instances of that single object/event, those which have been verified and those which are yet possible.

The generalizations, of both forms, that give rise to perceptive types or perceptive concepts are particularly important to every day knowledge because they allow for the identification, recognition and thereby the classification of the objects of the world and for their groupings into a class consisting of different objects in the case of *e pluribus generalizations* and of different perceptive instances of the same object in the case of *ex uno generalizations*.

4. Conceptual framework

The classical theory of concepts, as we have seen, states that the classes of objects are defined by necessary and sufficient criteria of belonging. All members of a class should therefore possess an equal degree of belonging to the class, that is, they should have all the attributes that typify that class and make up the necessary and sufficient conditions for membership. Actually, that may hold for some artificial and conventional concepts, but not for most empirical concepts. Indeed, classes are not logical entities defined by a set of necessary and sufficient conditions, but are determined by a set of identitive attributes most frequent (in the statistical sense) in members of the same class; from this set it is possible to obtain a typical exemplar by abstraction, one that does not possess the average value the highest number of characteristics shared by members of the class as expected in the *prototype theory* (Rosch 1977), but the identitive attributes common (most frequent) to those members.

With regards to the attributive structure of empirical concepts, it is essential to underscore that contrary to what many scholars maintain, an empirical concept does not consist solely of attributes referable to physical properties of objects, but also of other attributes assigned to objects by the individuals who experience them. For this reason we will be considering empirical concepts as *conceptual frameworks*, within which it is possible to distinguish two structures: a) an *identitive attributive nucleus* (*IAN*) referred to a class of objects or to different perceptive instances of the same object and defined according to statistical criteria, which consists of an empirical generalization, be it *e pluribus* or *ex uno*, operating on the basis of information contained in the mental representations corresponding to specific phenomenic objects; b) a specifying attributes set (SAS) referable to single objects belonging to a class, or in denotative terms, the set of specifying attributes or characteristic predicates of given members of a class. Furthermore, the *IAN*, made up of *identitive attributes* (or predicates), represents the *intention* of the concept, i.e., the set of predicates that, from a connotative point of view, define it. The *SAS*, on the other hand, is the set of attributes specifically characterizing each single object of one class, and also possess one, some or all of the attributes of the *IAN*. For example, in the case of the concept 'table', the *IAN* is made up of identitive attributes (or predicates) like 'having a rigid panel which is parallel to the ground', and 'having one or more supporting legs', whereas the *SAS* unites all of the specifying attributes (or predicates) referable to single members of the class of the tables such as 'having drawers', 'having leaves', 'being foldable', etc.

Therefore, we consider concepts as a *conceptual framework*s (*CONF*), which is represented by a five dimensional vector: *CONF* = <*IAN*, *SAS*, *SIG*, *SEM*, *NOM*>. These dimensions, examined in the sections below, correlate with one another and allow for a concept, according to precise modalities and a determinate state of mind, to perform the function of conceptual operator capable of triggering the processes of classification of the objects and subsequently to trigger different thoughts or actions.

At this point, it is necessary to formulate an analytical description of each dimension of the vector *CONF*.

4.1. The Identitive Attributive Nucleus (IAN)

The *IAN* is the result of an empirical generalization, that might be, as previously stated, either *e pluribus* or *ex uno*.

In the first case, the attributes of the *IAN* are individuated on the basis of their empirical statistical frequency as revealed in different objects, namely, they are most often found in perceptive instances of objects and, by induction, may be assigned to objects experienced at a later time.

These identitive attributes perceived are assigned both a series of *possible values*, that is, the range of possible values of an attribute, and a value of *default*, which is the value given to an attribute in a typical exemplar. For example, in the concept 'table', the attribute 'number of legs' might be '4' — the typical table has 4 legs — while the possible values are many: 1, 2, 3, 4, 5, etc. Analogously a human being, who generally has two eyes (the default value being '2'), may, due to a physical accident or to a genetic abnormality, have only one or even none (possible values 1 and 0). Therefore, the set of *default* values configures the *IAN*, that is, describes an object that can be immediately recognized as belonging to a given class.

Let us consider, for example, the class 'crows': an identitive perceptive attribute of this class is indisputably represented by 'blackness', the color whose high frequency characterizes the plumage of single crows. However, a white colored passerine with all of the identitive attributes of corvids, though lacking the identitive attribute of 'blackness', can regardless be considered a crow, even if the attribute of 'whiteness' found in that exemplar has a low statistical occurrence within the class of 'crow' and furthermore, can be considered a specifying attribute to that corvid. Thus, this bird might be placed by all rights in the class of corvids, without considering further its color (white) as an identitive attribute of the class it belongs to, but rather as a specifying attribute (a *SAS*'s attribute).

The same can be said with regards to a human being who due to an accident or genetic abnormality is without hands. In this case, the perceptive instance refers to an individual who possesses all of the other identitive attributes of the class of 'human beings' thereby rendering the absence of the identitive attribute 'hands' irrelevant from the point of view of statistical occurrence, because the coexistence of other identitive attributes legitimize their classification among the members of the class of 'human beings'.

In the second case, that of *ex uno* generalizations, an analogous process develops, but in reference to a single object which, as we have seen, has been exposed to several perceptive instances. It follows then that in the *ex uno* generalization, the frequency of the attributes refers to the perceptive instances of the same object.

It is therefore possible to define the IAN as the set of identitive attributes empirically checked on the basis of their statistical frequency with respect to a sample of objects or to the identitive attributes checked with the highest frequency related to different perceptive instances of the same object.

4.2. The Specifying Attributive Set (SAS)

The *SAS* is the set of perceptive specifying attributes (added to the identitive ones) pertaining to the single members of a class or of the single perceptive instances of the same object. In other words, in the case of a *e pluribus* generalization it is the set of all of the predicable attributes of the objects belonging to the same class (these objects at same time share one, some or all the predicates of the *IAN*) or to different perceptive instances of single objects in the case of a *ex uno* generalization.

In the case, for example, of the class of chairs, the *IAN* would be made of identitive attributes such as 'having a seat', 'having four legs', 'having a back', to which specifying attributes can be associated such as 'being foldable', 'having arms' etc., which make up the *SAS*.

Analogously, the *SAS* can also refer to many perceptive instances of one single object: in such a case, it is made up of the attributes of the experienced object in several perceptive instances (see Section 3.2.1). Furthermore, if we take into consideration 'my watch', the corresponding *SAS* will be made up of all the specifying attributes found in each perceptive instance of 'my watch' among which 'my watch this morning', 'my watch on my desk', 'my watch with the glass broken', etc.

4.3. Significances (SIG)

SIG consist of the different and possible significances assigned to the objects of one class of objects or a class formed by different perceptive instances of the same object. For example, our concept of 'watch' is not constituted by its features alone (the image of the watch as structurally isomorphic to experienced watches), but includes many significances that we have assigned to the watch that we use, that is, for example to 'our watch, in relation to a particular moment in our existence in which we bought it' or 'our watch with the sentimental significance referred to the person gave it to us', etc. Thus, as an example, we can assign many significances to the watch that we own: the significance of 'mine' which distinguishes it from other watches, making it unique with respect to the entire class to which it belongs; 'a gift from the woman I love', which bears an emotional–existential content to it referring to the giver; 'precious' or 'beautiful', which define the specific qualities of that particular watch, both from a material or aesthetic point of view and from an emotional one (precious because it is made of gold or is from a loved one).

4.4. The Reference (SEM)

SEM is a concept's reference to a class of objects or to one object. Thus, the concept 'cat' is also made up of its reference, that is, of all of the animals that possess the identitive attributes pertaining to the class of cats, and as such are classified as cats or even by a single cat that has been experienced repeatedly as would be true of the concept 'my cat' resulting in an *ex uno* generalization.

4.5. The name (NOM)

A 'name' as a linguistic expression is assigned to every concept thereby allowing for the nominability of the class of objects or the single object to which the concept refers. The name 'cat', for example, indicates the concept which refers to a specific class of animals.

The *NOM* brings the related concept into awareness and as such can be used for different purposes: to formulate different cognitive functions, to evoke emotions or mood states, to recall other concepts linked to the sharing of one or more attributes.

5. Visual concepts

In this section we will briefly consider the structure of what we call visual concepts, because in the human perceptive experience these are the most common and most relevant, and here we are referring in particular to the visual concepts formulated by *e pluribus* generalization. Analogously, we can analyze the visual concepts derived from *ex uno* generalizations.

The notable relevance of visual concepts has a neurophysiologic basis. The part of the CNS specialized in the elaboration of the visual stimuli (the primary and secondary visual areas in the occipital lobes) is more vast and complex than the cortical areas which elaborate information coming from the other sense organs.

Empirical concepts are the cognitive result of constant and intense perceptive activity of humans. Thus, there are empirical concepts of a tactile origin such as 'smooth' or 'soft', others that are olfactory such as 'stench' or 'fragrant', others that are auditory like 'sound' or 'noise', others still that are gustatory such as 'sweet' or 'salty'. Finally, there are concepts derived from the visual perception which represent the greater portion of empirical concepts and are made up of *figural* attributes (but also of *non–figural* attributes) predicable of an object. Indeed, a visual predicate refers to the figural attributes of the object (dimension, color etc.), while the non–figural attributes are related to the non figural structure of an object or to its functional attributes: 'being high' in reference to a mountain is a figural predicate, while 'being scrupulous' in reference to a human being is a non–figural predicate. Often in a concept's connotative definition both types of predicates are included, for example, in the case of the concept 'dog', 'airplane', 'house', even though concepts that do not have figural predicates are not rare, like the concept of 'tasty', or 'salty'. For example, to a dog one could predicate figural attributes like 'quadruped', or 'having a tail', and non–figural attributes like 'being loyal' or 'being domestic'.

At any rate, in daily life we rely heavily both on visual concepts which refer to single objects like 'my car', and on classes of objects like 'works of visual art', which are predominately figural. These concepts distinguish themselves from non–visual concepts in that figural attributes are those that define them and make up the identitive attributive nucleus which individuates them and distinguishes between them. Hence, their *IAN* is mainly made up of figural attributes. Visual concepts consist of identitive figural attributes which can be applied to a class of objects (even if the class consists of only one object). On the one hand, they can be *ex uno* and *e pluribus* visual generalizations, and on the other they can be formulated independently of visual generalization based on visual perceptions.

In the first case a visual concept derives from a visual generalization, which can be codified in figural or non–figural language and is stored without any further direct reference to cases which may have allowed for its formulation. As such, visual concepts play a role of figural types which refer to a class in accordance with the principle of *cognitive economy*, indispensable for a correct and efficacious functioning of the mind.

In the second case, there are visual concepts which are independent of generalizations and have no reference to cases or occurrences and their formulation derives from the use of visual material stored or formulated *ex novo*, although it utilizes material which has been stored previously. This is the case, for example, with visual concepts from the mythological matrix such as the unicorn, phoenix, or sphinx, which utilize visual perceptive material derived from experience referring to animals such as the horse and the lion albeit re–elaborated by our mind or through the visual perception of illustrations or photos.

Therefore, a visual concept consists of a set of entirely or primarily figural attributes and originates from the elaboration of information both visual and stored in the long term memory as schemes to be recalled sometimes, though not always, consciously in the moment of recognizing an object and assigning it to the class to which it belongs.

In ordinary thought visual concepts are rarely made up entirely of figural attributes because the objects are often perceived in a *perceptive space* that is broad and 'contaminated' by non–visual attributes. A "neutral" space, the "pure" background upon which objects can be seized and perceived by the subject in their empirical actuality exists only from a theoretical point of view.

The analysis of mental images (perceptive and non–perceptive) and of the processes involved aids in revealing the structure of empirical concepts and the specific dimensions (*IAN*, *SAS*, *SIG*, *SEM*, *NOM*) much more clearly, and consequently, in particular, the identitive and specifying attributes.

Indeed, as is clear from current neurophysiologic research on vision, after having been elaborated by visual apparatuses, including the areas of the visual cortex, visual stimuli are elaborated with information coming from different cortical and non–cortical areas and thus the image which gets formulated is the result of different and complex elaborations which involve cortical and non–cortical information. Hence, images cannot be reduced to the mere elaboration of information coming from visual stimuli (Bianca 2009:109–158).

As has already been shown, at the same time different visual mental representations are further elaborated in the formulation of generalizations (i.e. sharing predicates related to different representations), which are visual concepts whose *IAN*, contrary to other concepts, is made up primarily of figural attributes. The analysis of visual concepts, therefore, can corroborate our thesis on the structure of empirical concepts formulated in the previous sections.

Conclusive remarks

This article has examined the nature of empirical concepts, describing their structure through the formulation of a theoretical model based on statistical criteria which, if researched further and verified, could indicate a new perspective of study and analysis of the formation of conceptual structures based on perceptive instances.

The formation of empirical concepts is a fundamental process for knowing and describing the physical world, which is useful for "ordering" our experience, distinguishing one object from another, recognizing the respective differences and for acting in our world. In human beings, evidently, empirical concepts possess a complex structure because they are influenced by the cultural environment and their formation includes information which comes from different cortical, sub–cortical, and non–cortical areas. To this end, we have examined the structure of empirical concepts not only from the point of view of their perceptive content, but also taking into consideration other information of a non–perceptive nature (for example, the different significances that a subject can attribute to one or more objects belonging to a class) generally overlooked in studies and researches carried out to date both in the field of philosophy and cognitive psychology.

Empirical concepts are undoubtedly the means for formulating an awareness of the physical world. They speak of this world and allow for its description by virtue of their perceptive foundation and hence lend a reliable and adequate (as discussed) empirical understanding whose end is to live and thrive within it.

In the light of these considerations a model was formulated in which concepts are considered in their complexity as *conceptual framework* and each one represented by a five–dimensional vector in which, in addition to the merely perceptive contents (*the identitive and specifying perceptive attributes*), non–perceptive contents (semantic references and various significances assigned to concepts by single individuals) are analyzed. This model does not consider empirical concepts as the simple result of a generalization conducted on the basis of different perceptive instances, but highlights the relevance of non–perceptive contents involved in their formation process.

The model outlined in this paper tries to overcome the contrast between perceptive and non–perceptive mental processes since many empirical concepts contain non–perceptive attributes resulting from an autonomous neuromental elaboration carried out by every single individual in relation to her/his experience with the information contained in the different neuromental areas; attributes assigned on the basis of statistical frequency and on the two types of generalization; therefore, contrary to what is generally claimed, there are not only concepts that refer to a class of objects (*e pluribus* generalizations), but also concepts that refer to a class of perceptive instances related to single objects (*ex uno* generalizations).

Empirical concepts possess a complex informative structure such as that indicated in section 4, and cannot be reduced to the mere elaboration of information coming from perceived stimuli. This model requires further in–depth analysis, including experimental trials, in order to clarify aspects in this article that have not been dealt with or that have only been touched upon briefly.

References

Armstrong, S., Gleitman, L., Gleitman, H. (1983). What some concepts might not be. *Cognition*, 13, 263–308.

Bianca, M. L. (2005). *Rappresentazioni mentali e conoscenza. Un modello teorico–formale delle rappresentazioni mentali*, Milano, FrancoAngeli.

Bianca, M.L., Foglia, L., (2008). La cognizione figurale, FrancoAngeli.

Bianca, M.L., Lucia Foglia (2006). *Non–Perceptive Mental Image Generation: a Non–Linear Dynamic Framework*. Anthropology & Philosophy, 7, 28–63.

Bianca, M.L. (2009). *La mente immaginale. Immaginazione, immagini mentali, pensiero e pragmatica visuali*, Milano, FrancoAngeli.

Bianca, M.L., Piccari, P. (2007). Inherent Logic: Isotopic and Inherent Bonds in Argumentation. *Anthropology & Philosophy*, 8, 1–2, 9–31.

Frege, G. (1891a). Über das Trägheitsgesetz. Zeitschrift für Philosophie und philosophische Kritik, 98, 145–161, poi in Id., *Kleine Schriften*, ed. I. Angelelli, Olms, Darmstadt 1990^2, 113–124.

Frege, G. (1891b). Funktion und Begriff. *Vortrag gehalten in der Siztung vom 9. Januar 1891 der Jenaischen Gesellschaft für Medicin und Naturwissenschaft*, H. Pohle, Jena 1891, poi in Id., *Kleine Schriften*, ed. I. Angelelli, Olms, Darmsdat 1990^2, 125–142.

Frege, G. (1892). Über Begriff und Gegenstand. *Vierteljahrsschrift für wissenschaftliche Philosophie*, 16, 192–205, poi in Id., *Kleine Schriften*, ed. I. Angelelli, Olms, Darmstadt 1990^2, 167–178.

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100, 1892, 25–50, poi in Id., *Kleine Schriften*, ed. I. Angelelli, Olms, Darmstadt 1990^2,.143–162.

Gelman, S.A., Coley, J.D. (1991). Language and categorization: The acquisition of natural kind terms. In Gelman, S.A., Byrnes, J.P. (eds.). *Perspectives on Language and Thought. Interrelations in Development*, Cambridge, Cambridge University Press.

Girotto, V., Legrenzi P. (eds.). (1999). *Psicologia del pensiero*, Bologna, il Mulino.

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

Keil, F.C. (1989). *Concepts, Kinds, and Cognitive Development*, Cambridge (MA), The MIT Press.

Kneale, W.C. (1949), *Probability and Induction*, Oxford, Oxford University Press.
Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind,* Chicago, University of Chicago Press.
Lalumera, E. (2009). *Che cosa sono i concetti*, Roma–Bari, Laterza.
Laurence, S., Margolis, E. (2002). Concepts and Concept Analysis. *Philosophy and Phenomenological Research*, LXVII, 2, 253–282.
Machery, E. (2009). *Doing without Concepts*. New York, Oxford University Press.
Margolis, E. (1998). How to Acquire a Concept. *Mind & Language*, 13, 347–369.
Margolis E., Laurence, S. (1999). Concepts and Cognitive Science, in Idd. (eds.), *Concepts. Core Readings*, Cambridge (MA), The MIT Press, pp. 3–81.
Margolis E., Laurence, S. (2007). The Ontology of Concepts. Abstract Objects or Mental Representations? *Noûs*, 41, 561–593.
Murphy, G. L. (2002). *The Big Book of Concepts*, Cambridge (MA)–London, The MIT Press.
Murphy, G. L., Medin, D.L. (1985). The Role of Theories in Conceptual Coerhence. *Psychological Review*, 92, 289–316.
Prinz, J.J. (2002). *Furnishing the Mind. Concepts and Their Perceptual Basis*, Cambridge (MA), The MIT Press.
Rips, L. J. (1989). Similarity, tipicality, and categorization. In Vosniadou, S., Ortony, A. (eds.). *Similarity and Analogical Reasoning*, Cambridge, Cambridge University Press, 21–61.
Rosch, E. (1975a). Cognitive Reference Points. *Cognitive Psychology*, 7, 532–547.
Rosch, E. (1975b). Cognitive Representation of Semantic Categories. *Journal of Experimental Psychology: General*, 104, 192–233.
Rosch, E. (1977). Human Categorization, in Warren, N. (ed.) *Studies in Cross Cultural Psychology*, I, London, Academic Press.
Rosch, E. (1978), Principles of Categorization, in Rosch, E., Loyd, B.B. (eds.), *Cognition and Categorization*, Hillsdale (NJ), Erlbaum.
Rosch, E., Simpson, C., Miller, R.S. (1976). Structural Bases of Tipicality Effects. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 491–502.
Smith, E.E., Medin, D.L. (1981). *Categories and Concepts*, Cambridge (MA), Harvard University Press.

Mariano L. Bianca – Paolo Piccari
University of Siena
bianca@unisi.it – piccari@unisi.it

L'articolo è stato elaborato, discusso e rivisto congiuntamente dagli autori. In ogni caso, soltanto a fini accademici, è possibile attribuire le sezioni 3, 5 6 a Mariano Bianca e 1, 2 e 4 a Paolo Piccari.

Conjunctions of Social Categories considered from Different Points of View

JAMES A. HAMPTON, MARGARET DILLANE, LAURA OREN AND LOUISE WORGAN

Abstract Conjunctions of divergent social categories may elicit emergent attributes to render the composite concept more coherent. Following Kunda, Miller & Clare (1990) participants listed and rated attributes for people who belong to unexpected conjunctions of social categories. In order to explore the flexibility in such constructions, they were also asked to adopt the point of view of a person in one of the two categories. Experiment 1 found that when adopting the point of view of one constituent category, people tended to combine the concepts antagonistically, meaning that they attributed to members of the conjunction the more negative aspects of the opposing category. Experiment 2 showed that this polarizing effect was reduced when the point of view category was itself unusual. Strong gender stereotype differences were also found in the degree to which combinations were antagonistic. Female stereotypes as points of view generated a greater degree of integration in the conceptual combination.

Key words Conceptual combination, stereotypes, gender

The experiments described here stem from a research question originally raised by Osherson and Smith (1981) concerning conjunctions of fuzzy concepts. A fuzzy concept is one that defines a category of exemplars containing not only clear–cut examples, but also atypical and borderline examples. Zadeh (1965) introduced the notion of a fuzzy set to model the reference of fuzzy predicates such as "is tall" or "is red". A man's degree of membership in the category of tall men is some continuous function of his height, rather than simply being true above some value and false below it. Zadeh proposed that one could represent degree of set membership with a logical function *c* that could take continuous values between 0 (clearly not in the set) and 1 (clearly in the set).

Rosch and Mervis (1975) extended the idea of fuzziness to cover the extensions of nouns as well as adjectival modifiers. In their prototype theory they argued that concepts such as fruit or furniture could also be characterized by fuzzy sets. As evidence they showed that people can consistently rate the typicality of category members, and that this typicality variable affects a range of psychological measures of category processing. Hampton (1979) and McCloskey & Glucksberg (1978) showed that many of these noun categories were indeed fuzzy in that people were both in disagreement with each other, and inconsistent themselves in deciding whether borderline cases belonged in the category.

The question then arises of how such concepts could be combined by logical operations such as conjunction or disjunction. In his fuzzy set logic, Zadeh (1965) proposed two possible rules that could be applied to form the conjunction of fuzzy sets, both of which had the desirable property that if applied to all–or–none concepts, the traditional definition of conjunction would emerge. One rule was the minimum rule, whereby set membership in a conjunction was the minimum of the two constituent set memberships. The other rule was a product rule, which proposed conjunctive set membership to be the product of the two constituent set memberships. However Osherson & Smith (1981, 1982) argued persuasively that the application of fuzzy intersection rules to category conjunctions such as pet fish or striped apple was doomed to failure. In particular, Zadeh's two rules for conjunction of fuzzy sets entailed the inequality that no item could be a better member of a conjunction than it was of either of the two constituent categories. Yet intuition seemed to argue otherwise, and Smith & Osherson (1984) provided data to make the point that items were frequently considered more typical of a conjunctive concept than of a constituent. Thus a brown apple was more typical of the concept brown apple than of apple, and a guppy was argued to be more typical of the concept pet fish than of either the concept pet or the concept fish. Osherson and Smith (1982) were also able to present a logical demonstration that no function could be found that would successfully map constituent concept typicality onto conjunctive concept typicality for all pairs of concept categories.

It soon became apparent in the psychological literature that to account for such conjunctions, it was necessary to provide an intensional model for combining the attributes of each concept, rather than a truth–functional extensional model which would follow Zadeh's approach in defining membership in a conjunction as a function of the membership values for each constituent.

These models (Cohen & Murphy, 1983; Hampton, 1987, 1988; Smith, et al. 1988) aimed to explain the conjunction of fuzzy concepts by showing how the attributes of the two concepts are combined. In this way the impossibility of providing a function to map typicality and membership of constituents onto typicality and membership for conjunctions can be circumvented.

Hampton's composite prototype model for conjunctions (Hampton, 1987, 1988) showed how two category concepts, defined as prototype attribute specifications, could be combined into a composite prototype for the conjunction of the concepts. The model proposed that to form a conjunction A∩B of concepts A and B, the two lists of attributes that define A and B are initially combined into a single list. This first step predicts (a) that degree of typicality in a constituent (number of constituent attributes possessed) will predict degree of typicality in the conjunction, and (b) that if two concepts are combined for which one concept has a higher number of centrally important attributes than the other, then degree of membership in this constituent will dominate membership in the other. Both of these predictions were confirmed.

The model also proposed a second step in order to account for non–compositional effects in category conjunction. The two lists of attributes may contain mutually incompatible values – pets are warm and cuddly whereas fish are cold and slippery. Pet fish clearly can not be both. Attributes are thus lost from the composite prototype in order to render the conjunctive concept consistent. Experience with real world objects, and background theories of the world may both be involved in this stage.

Following this research with semantic categories, Kunda, Miller & Clare (1990) investigated the effects of combining concepts relating to human social categories (see also Hastie, Schroeder & Weber, 1990; Hutter & Crisp, 2005; Kunda & Thagard, 1996). For example, when faced with the concept of a Harvard educated carpenter, what do people generate as the expected composite prototype concept? It appeared from their research that people invoke a range of social theories and background knowledge to "explain" the novel combination, in keeping with an approach to conceptual combination advocated by Murphy (Cohen & Murphy, 1983; Murphy, 1988; Murphy & Medin, 1985). Attributes that were generated for the conjunction focussed on why a Harvard educated person should work as a carpenter, and what kind of carpentry would be expected to result from a Harvard education. Social categories therefore appear to be a rich area for investigating "non–monotonic" effects in category conjunction, and the experiments described here used similar conjunctions of social categories in order to explore this richness further.

The manipulation introduced in the present research was based on a series of investigations by Barsalou & Sewell (1984) into the effects on category structure of asking people to adopt different points of view. Participants in their studies were asked to rate and to rank the typicality of objects in various categories, while adopting different points of view. For example participants rated the typicality of vehicles from the point of view of suburban housewives or redneck farmers. The results showed that point of view could completely change the typicality ranking of category exemplars, and that groups of students were remarkably consistent in adopting the points of view of other groups. When undergraduate, postgraduate and faculty members' points of views were compared, there was even a close agreement between the consensus viewpoint of one group of another's category structure and the other's own viewpoint as they themselves expressed it.

Given that taking a point of view has such a marked effect on conceptualization, we decided to use the manipulation in the current experiments in an attempt to see how adopting the point of view of belonging to a particular social category would affect the degree to which attributes of the category are inherited in a conjunction. Whereas Barsalou and Sewell (1984) measured typicality of category members, there has been no previous study looking at the impact of adopting a point of view on the intensional attributes considered to be true of the class. It was expected that combining a manipulation of point of view with the study of attribute inheritance in category conjunctions would offer insights into both how conjunctions are formed, and how point of view affects category representations. In particular, if non–monotonic effects in conceptual combination are the result of an appeal to background theory in the way that Murphy and Kunda et al. propose, then a change in point of view should produce a marked change in the characterization of the conjunctive concept, and hence should affect the kind of emergent attributes that may be found.

Experiment 1

Method

Participants. Participants were students and other adults associated with City University London. All were volunteers and native speakers of English and familiar with British society. Thirty participants generated attributes in stage 1, and 96 participants rated attributes in stage 2. Participants were allocated to conditions at random regardless of gender. There were approximately equal numbers of males and females. Those rating attributes were paid £5 for their participation.

Materials. Four pairs of social categories were chosen, relevant to a British population. They were: Conservative Party Supporter/Trade Unionist, Socialist/ Stockbroker, Oxford Graduate/Factory Worker, and Rugby Player/Man Who Knits.

The concepts were arranged in pairs to represent antagonistic categories–members of one could belong to the other but would be unlikely to. As in the study by Kunda et al. (1990), there was an element of improbability that someone should belong to both categories, as it is for these combinations that people are most likely to generate emergent attributes for the conjunction. In addition the categories represent groups that according to stereotypes may be considered antagonistic in their attitudes to each other.

Procedure. In the first phase of the experiment, participants generated attributes to both constituent and conjunctive categories. In doing so they were always asked to adopt the point of view of one or other of the constituent categories. They were asked to list characteristics of a person in answer to the question: "What would you expect someone with the given point of view to say about the person to be described".

In the second phase, lists of attributes generated in the first phase were combined for each pair of categories, and a new sample of participants were asked to rate how appropriate each attribute was for each category and each conjunction, from a given point of view. Instructions were as follows:

'This study is about the views people have of each other. On each page you will be given a type of person and a point of view from which they may be described together with a set of characteristics which might be used to describe them. Your task is to adopt the given point of view and then rate the characteristics for their appropriateness."

An example was given. The rating scale was a five point scale labeled "Highly Inappropriate, Inappropriate, Neutral, Appropriate, and Highly Appropriate."

Design. There were six groups of participants for the attribute generation task, with five participants in each. Two groups generated attributes to individual constituent categories, and four generated attributes to conjunctions. Each participant in the two constitutent groups completed a booklet with eight pages, each listing one of the eight constituents, with point of view balanced across the two groups. Thus each participant described each constituent once only. Order of constituents was balanced across participants, and there was a constant lag of 4 between the first and second category from each pair. Participants in the four conjunction groups generated attributes for different versions of the four conjunctions, from different points of view. A given pair could be presented as a relative clause construction in either order (A Conservative party supporter who is a Trade Unionist or A Trade Unionist who is a Conservative party supporter), and these two could be rated either from a Conservative party supporter or a Trade Unionist point of view. These four sets of ratings were distributed across the four groups. Order of rating the four conjunctions was randomized for each participant.

From the attributes listed by the six groups of participants, a master list of attributes was drawn up for each concept pair. The list included the four most frequently listed attributes from each of the eight lists provided by the six participant groups — that is from each constituent and from each order of their conjunction taken from each point of view. Where there was overlap, an attribute was only listed once. Because of overlap, the final lists contained between 24 and 29 attributes, ordered alphabetically. These master lists were then used in stage two of the experiment.

The same design was used for attribute rating as for the generation task, in order to obtain ratings for the master lists for each constituent and each order of the conjunctions from each point of view. As before, ratings for constituents were doubled up so that each participant rated 8 lists, whereas participants in the conjunction rating groups just rated 4 lists. Participants were asked to adopt the particular point of view and then to rate the "appropriateness" of each attribute for the category on a five point scale from Highly Appropriate through Neutral to Highly Inappropriate. Sixteen participants acted in each of the six groups, so that there were sixteen ratings of each attribute in a list for each constituent from each point of view and for each conjunction order from each point of view.

Results

The results are presented in two sections. The first analysis used a correlational approach to consider two issues: first, the prediction of the composite prototype model that importance for a constituent should predict importance for the conjunction, and second the effect that the manipulation of point of view had on the relative importance of each constituent in determining the conjunction. The second analysis dichotomized the scales so that each attribute was coded as True or False of each concept. (For ease of exposition the rating scale of "appropriateness of description" is treated as a measure of the truth of the attribute as applied to the concept). Cross tabulation was then used to examine the patterns of attribute inheritance. Three questions were addressed: (a) to what extent did the attributes that were judged true of a conjunction correspond to the set union of the constituent attributes, (b) to what extent were there emergent attributes, true of the conjunction but not of either constituent, and (c) how did changing the point of view affect the attributes which were inherited?

Correlational Analysis. The correlational analysis required the calculation of a mean appropriateness rating for each attribute for each of the constituents and each conjunction, taken from each point of view. The five point scale was coded numerically with a +2 for Highly Appropriate, 0 for neutral and –2 for Highly Inappropriate. There were eight mean ratings to be calculated for each attribute, corresponding to four categories — the two constituent concepts, and the two conjunctions (depending on the order of concepts) — considered from two possible points of view. Since a correlational analysis was intended, it was important to establish reliable measures for each of the scales. Before calculating mean values for the ratings, a reliability analysis was done within each attribute list, by considering the correlation of each participant's ratings across the list with the total ratings summed across the remaining participants in the group. Participants with negative correlations with the group total for a particular attribute list were excluded from calculation of the mean for that list (between 0 and 5 out of 16 participants were excluded per list).[1]

Final mean reliability for the conjunctions was .88 (Cronbach's alpha) with a range from .75 to .96. For constituent concept ratings, final mean reliability was .95 with a range from 89 to .98. The familiar stereotypes were therefore more reliably rated than the unfamiliar (and improbable) conceptual combinations.

Within each of the four concept pairs, correlations were calculated across attributes between the eight mean ratings: constituent *A*, constitu-

1. The justification for eliminating participants at this stage is that the study is concerned with finding an accurate measure of the prevailing stereotype within the population sampled. Some individuals had different views for some concepts, and including them in the mean would have obscured the results obtained by adding noise to the measures.

ent *B*, *A* who are *B*, and *B* who are *A*, each as rated from point of view *A* or from point of view *B*.

Correlations for the same concept rated from opposing points of view are shown in Table 1. For example the appropriateness of applying the list of attributes to Socialists as judged from the Socialist point of view correlated at –0.02 with the appropriatness of the attributes as applied to Socialists from the Stock Broker point of view. Thus if point of view had no effect at all, correlations should be high and approach the mean reliability of the two measures. The effect of changing point of view varied across concepts. In many cases the manipulation of point of view had a very strong effect (zero correlation or even a significantly negative correlation), while in other cases the manipulation was relatively weak (showing a strong positive correlation between the two points of view). For example, participants imagining themselves to be Conservative party supporters or trade unionists shared quite a similar view of a Conservative party supporter who is a trade unionist ($r = 0.68$), although they held quite different views about each other ($r = -.15$ and $-.27$). On the other hand those taking the point of view of rugby players and men who knit shared a similar view of rugby players ($r = .76$), but significantly opposed views about men who knit who are rugby players ($r = -.49$). Effects of point of view clearly interacted with the semantic content of the categories in question, since no consistent pattern was seen across categories here. However it can be concluded that point of view was having a powerful effect on most of the categories, with only 4 of the 16 correlations showing a significant positive relation between the two points of view.

Concept A	Concept B	A	B	A∧B	B∧A
Conservative Supporter	Trade Unionist	–.15	–.27	.68*	.32
Socialist	Stock Broker	–.02	.43	.25	.38
Oxford Graduate	Factory Worker	.71*	.35	.55*	.42
Rugby Player	Man who Knits	.76*	–.01	–.25	–.49*

Table 1. Correlations between Different Points of View of the Same Category, in Experiment 1. *Note*: A∧B = "Concept A who is Concept B".

Table 2 shows for each point of view the correlations of each constituent with each conjunction, and the intercorrelation of the constituents. For example, attribute ratings for the constitutent Socialist correlated at –.683

with those for Socialist who is a Stockbroker, from the Socialist point of view, and +.303 with the same conjunction from the point of view of a Stockbroker. For the conjunction Stockbroker who is a Socialist, the corresponding correlations were –.502 and +.418. The final column shows the correlation between the two constituents. In this case attribute ratings for Socialist correlated with those for Stockbroker at –.654 from the Socialist point of view, and at -.821 from the Stockbroker point of view.

On the basis of earlier research (Hampton, 1987) importance of an attribute for a conjunction would normally be expected to correlate positively with importance for each constituent. For example an attribute such as "is competitive" would be as important for the conjunction "sports that are games", as it was for sports or games alone. However for 7 of the 8 points of view, a radically different pattern was seen here — a strong negative relation between the two constituents, a positive correlation of the conjunction with the other's point of view and a negative correlation with one's own point of view.

Point of View	Conjunction Order	Constituent A with Conjunction	Constituent B with Conjunction	A with B
Conservative (A)	AB	–.794	.813	–.759
	BA	–.310	.537*	
Trade Unionist (B)	AB	.694*	–.646*	–.774*
	BA	.864*	–.863*	
Socialist (A)	AB	–.683*	.776*	–.654*
	BA	–.502*	.649*	
Stock Broker (B)	AB	.303	–.137	–.821*
	BA	.418*	–.293	
Oxford Graduate (A)	AB	–.273	.370*	–.736*
	BA	–.205	.326	
Factory Worker (B)	AB	.287	–.478*	–.406*
	BA	.625*	–.305	
Rugby Player (A)	AB	–.643*	.859*	–.866*
	BA	–.799*	.940*	
Man who Knits (B)	AB	.373*	.761*	–.177
	BA	.423*	.761*	

Table 2. Correlations between Constituents and Conjunctions in Experiment 1. *Note*: Conservative = Conservative Party Supporter. * significant at 0.05.

For 7 of the 8 points of view shown in Table 2, a radically different pattern was seen from that expected. A strong negative relation was found between the two constituents, together with a positive correlation of the conjunction with the other's point of view and a negative correlation with one's own point of view. The positive correlations were generally the stronger of the two, so that when the interconstituent correlation was partialled out, the correlation with one's own point of view was often near zero. That is to say there was no correspondence between the attributes considered true of the category adopted as the point of view and the attributes considered true of the conjunction from that point of view. The partial correlation for the constituent that was not the point of view adopted (the Other constituent) was always found to be positive, and was significantly greater than zero in 11 of the 16 conjunctions. The constituent that was the point of view adopted (the Own constituent) had significantly negative partial correlations for 3 conjunctions and significantly positive partial correlations in 2 (both of which involved the Man who knits point of view). These results were confirmed in a set of regression analyses shown in Table 3.

Point of View	**Beta weights**		**Adjusted R^2**	**Alpha^2**	**Percent explained**
Constituent	**Own**	**Other**			
Conservative	–.135	.645*	.532	.671	79%
Trade Unionist	–.393	.501*	.689	.880	78%
Socialist	–.229	.583*	.530	.891	59%
Stock Broker	.266	.610	.108	.687	16%
Oxford Graduate	.040	.386	.049	.893	5%
Factory Worker	–.267	.359	.212	.832	25%
Rugby Player	.213	1.111*	.858	.902	95%
Man who Knits	.873*	.563*	.896	.935	96%

Table 3. Regression Statistics for Predicting Importance for a Conjunction from Importance for each Constituent, within a Particular Point of View, Collapsed across Order of Conjunctions. *Note*: Own = Constituent category whose point of view was adopted, Other = Constituent category whose point of view was not adopted. Alpha = reliability of scale. Percent explained = percent of reliable variance accounted for (R^2/α^2). * significant at .05.

Regression equations were calculated for each point of view, predicting importance for the conjunction from importance for each constituent, all considered from the same point of view. Since the regression statistics were generally similar for each conjunction order, the equations were calculated using importance averaged across the two conjunction orders. The results are shown in Table 3, together with the adjusted R squared, which is the proportion of total variance explained, the squared reliability of the conjunction ratings which indicates the proportion of reliable variance there was to explain, and the former expressed as a percentage of the latter.

There was wide variation in the degree of fit of the regression models. For the Rugby Player/Man who Knits combinations, practically all the reliable variance in importance for the conjunction could be predicted from constituent importance. For the Oxford Graduate/Factor Worker example, neither of the equations reached a significant level of prediction. The degree of fit also sometimes depended on point of view — for example the socialist point of view gave a better fit (59% explained) than the stockbroker's point of view (16% explained).

For the constituent which was not the point of view (the column labeled Other in Table 3), the beta weights were clearly positive, with a mean of .575, and with 5 of the 8 equations showing significance. This pattern is consistent with earlier results on attribute inheritance (Hampton, 1987, Kunda et al., 1990). For the constituent which corresponded to the point of view adopted, (labeled Own in Table 3), across the first three category pairs the trend was for the point of view constituent to have little positive predictive value (mean = –.109). A negative regression weight implies that across the attribute list, the more an attribute was true of the point of view constituent, then the less it was considered true of the conjunction — a result that is clearly at odds with any existing theories of attribute inheritance.

The exception to this pattern was the rugby player who is a man who knits. For this combination when taking the man who knits point of view, the more normal pattern of positive regression weights for each constituent was found. The man who knits was the only constituent to positively predict importance for the conjunction when it was the point of view category.

To recap the results, there was a consistent pattern across the first seven of the eight points of view categories. When people considered a conjunction from the point of view of one of the constituents, they tended to see the attributes of the conjunction as most reflecting those of the other group. For example how true an attribute was seen to be of a socialist who is a stockbroker depended on which point of view one took. From the socialist point of view it depended on how true the attribute was of a stockbroker, while for the stockbroker point of view it was the degree to which the attribute correctly described a socialist that was influential. The one exception was the point of view of the Man Who Knits, where the two constituents were integrated into the combination in a positive fashion.

Attribute inheritance. In order to provide a criterion for investigating attribute inheritance, each attribute was rescored for each scale as simply True or False. A cut–off point of 3.5 on the numerical scale was used to dichotomize the data, corresponding to half way between the Neutral and the Appropriate points on the scale. To simplify the analysis, the two orders of the conjunction were combined for this purpose.

Because the social categories were generally antagonistic, it was felt useful for the analysis to separate out attributes which were considered positively valued from those which were considered negatively valued from a particular point of view. We can then see whether it is primarily positively or negatively valued attributes that are inherited. Two new groups of participants (10 participants in each) were given the attribute lists for each pair of concepts and rated each attribute on a 7 point scale from "very good" to "very bad", according to whether the attribute was something that was generally a good or a bad thing to be from a particular point of view. One group was asked to adopt one constituent as the point of view in making the judgments, and the other group adopted the other point of view. Means were calculated for each attribute from each point of view, and the scales were dichotomized around the neutral mid–point of the scale to give categories of positively and negatively valued attributes. Point of view was thus kept constant in analyzing both attribute inheritance, and the evaluation of the attributes.

	Positively evaluated		Negatively evaluated	
	True of conjunction	Not true of conjunction	True of conjunction	Not true of conjunction
True of Neither	4	26	18	15
True of Own	11	54	0	0
True of Other	7	6	48	19
True of Both	2	0	0	0

Table 4. Attribute inheritance for positively and negatively evaluated attributes true of each constituent. *Note*: Own = point of view adopted. Other = point of view not adopted.

The inheritance analysis results are shown in Table 4. Of those attributes rated true of neither constituent, 4 of the 30 positive attributes (13%), and 18 of the 33 negative attributes (55%) were nonetheless considered true of the conjunction. In other words there were 22 "emergent" attributes, across the 8 conjunctions. These are listed in Table 5.

Of the attributes true of just one constituent, if that constituent was the point of view adopted (True of Own in the table), then it was always a positive attribute. Of 65 such attributes, only 11 (17%) were inherited by the conjunction. For the attributes that were only true of the Other constituent (i.e. that which was not adopted as the point of view), most were negative (67 out of 80) and the inheritance rate was 72% for the negative, and 54% for the positive.

The poor fit in some of the regression models can now be explained here in terms of the number of emergent attributes seen. For example Table 5 shows that there were many emergent attributes for the Socialist/Stockbroker and Oxford graduate/Factory worker combinations, for which the fit had been poor. By contrast, the Rugby player who is a Man who knits, which had a well fitting regression model, showed very few emergent attributes — indeed none at all for the Man who knits point of view.

Category A	Category B	Point of View	Emergent Attribute
Conservative	Trade Unionist	Conservative	Traitor
Conservative	Trade Unionist	Trade Unionist	Traitor Confused Strange
Socialist	Stockbroker	Socialist	Champagne Socialist Traitor Unconventional Unrealistic
Socialist	Stockbroker	Stockbroker	Champagne Socialist Not a true socialist Unconventional
Oxford Graduate	Factory Worker	Oxford Graduate	Disaffected Lazy Possible mental breakdown Something wrong with him Unconventional
Oxford Graduate	Factory Worker	Factory Worker	Failure Something wrong with him Unconventional Under–achiever

Category A	Category B	Point of View	Emergent Attribute
Rugby Player	Man who knits	Rugby Player	Confused Relaxed
Rugby Player	Man who knits	Man who knits	—

Table 5. Emergent Attributes in Experiment 1. *Note*: Conservative = Conservative Party Supporter.

Discussion

The composite prototype model (Hampton, 1987, 1988) predicts that attributes of each constituent should normally be inherited by the conjunction. How well have the results supported this prediction? First, from the correlational analysis, there was only one category pair where the model provided a good fit — the Rugby player who is a Man who knits, considered from the point of view of the Man who knits. In each order of the conjunction for this point of view, the regression model explained almost all the reliable variance, and both constituents had significant positive beta weights. For the remaining concepts, the pattern of results failed to support the model — regression weights and partial correlations for the Own constituent were non–significant or even significantly negative, and there was considerable inheritance failure for the attributes true of the Own point of view constituent. Note that no existing model of conjunction formation would predict negative weights in these regressions. Kunda et al. (1990) found positive weights in their social category combinations. However in some cases the fit in their data was also very poor with less than 10% of the reliable variance explained, suggesting that importance for the conjunction was independent of any constituent importance.

The attribute inheritance pattern shows perhaps more clearly what is occurring in the current data. With the exception of the point of view of the man who knits, 95% of inherited attributes were inherited from the other's consistuent category, and 84% were unfavorable. There were also 22 (28%) emergent attributes for the conjunctions — again mostly negatively valued.

The emergent attributes (see Table 5) were similar to those identified by Kunda et al. (1990). Some simply reflected the surprising nature of the combination — for example *unconventional* or *confused*. Others offered a more explanatory account, as in *lazy*, *under–achiever* or *traitor*. Others might well be considered emergent, although they were not rated as such. For example *hypocrite* was generated in the lists for both Conservative party supporter/Trade unionist and Socialist/Stockbroker, and was rated as true of all four versions of the conjunctions. That this attribute did not appear as emergent was the result of the fact that each point of view saw the other category as possessing the attribute already!

One purpose in conducting the study was to investigate whether the types of "theory–based" reasoning used to account for unusual combinations of social categories would vary as a function of point of view. Evidence for this effect was relatively weak, but there was some observed difference for the Oxford graduate who is a factory worker category (our equivalent of Kunda et al.'s Harvard educated carpenter). Causal reasoning in this combination generated emergent attributes to explain how the person became or continues to be in this situation. Furthermore whereas from the Oxford graduate point of view the conjunction was seen as disaffected with the establishment, lazy or having suffered a mental breakdown, from the factory worker point of view there was more focus on failure and under–achievement.

Point of view was just as powerful a manipulation affecting attribute ratings, as Barsalou and Sewell (1984) had found with typicality judgments. However both the effect of point of view, and the process of attribute inheritance itself appear to vary widely across different semantic domains. The Conservative party supporter who is a trade unionist was an example of a very direct contradiction in the minds of the participants (although Mrs Thatcher's election victories in Britain in the 1980s were in part due to the votes of disaffected trade union members). In this antagonistic situation there was a tendency for attributes to become polarized in value — one's own side is good and the other is bad — and for the bad to drive out the good in the case of the conjunction. Thus the more true an attribute was considered to be of the Own point of view category (good) the less true it was considered to be of the conjunction.

Perhaps the most interesting of the differences between categories occurred for the Rugby player who is a Man who knits. From the rugby player's point of view, a man who plays rugby and who knits is basically like any man who knits. The importance of an attribute for a rugby player made no significant contribution to the regression equation, and the inheritance pattern showed the conjunction to be composed largely of the (negatively valued) attributes of the Man who knits. However taking the man who knits point of view, a picture of the tolerant "New Man" emerges. Both constituents were positively weighted in the regression, and the conjunction was composed of positively valued attributes from both constituents. It was the only one of the points of view in which the conjunction strongly reflected the good attributes of the Own point of view constituent. The conclusion is therefore that antagonistic categories are not all alike. We need a clearer understanding of the different kinds of unusual combinations that can occur if the process of conceptual combination is to be understood. The second experiment was designed to test one hypothesis of why some combinations are antagonistic and others are not.

Experiment 2

The first experiment produced two quite distinct patterns of conceptual combination when the point of view of one or the other social groups was adopted. In seven of the social categories, the adoption of a point of view led to antagonistic combinations, by which someone in the conjunction of two inconsistent stereotype categories was judged to be most like the "Other" category. In one of the categories however — the Man who knits — the combination was much more in line with the findings for natural categories like sports, games, pets or birds (Hampton, 1987), in that attributes were inherited from both constituent categories, and with no antagonism. There may be different post hoc hypotheses concerning this effect. For example, it was the only category which incorporated a relative clause construction in its description. A man who knits is also different from the other categories in that it is itself an unusual category, involving as it does a stereotype which is inconsistent with the gender of the target (knitting in England is most commonly seen as a stereotypically female activity). The hypothesis tested in the second experiment was that the adoption of a gender consistent point of view (a woman who knits, or a man who plays rugby) would result in a more antagonistic conceptual combination than the adoption of a gender inconsistent point of view (a man who knits or a woman who plays rugby). While this hypothesis was generated purely as a post hoc account of the results of Experiment 1, it can also be justified theoretically. Gender consistent points of view are likely to refer to social groups that are "normal" or generally accepted by society. Such groups (such as men who play rugby) may be imagined to be less willing to accept the eccentricity of those within them who also belong to gender inconsistent social groups (men who play rugby and also knit). On the other hand, someone belonging to a gender inconsistent group (such as a man who knits) has broken out of the stereotypical mold for their gender, and so may be imagined to be much more tolerant of people in the group who also belong to gender consistent groups (men who knit and also play rugby).

Method

Participants. Participants were students and other adults associated with City University London, from a range of different social backgrounds and occupations. All were volunteers and native speakers of English and familiar with British society. There ages ranged from 16 to 61 years. Eighty–four participants (51 females and 33 males) generated attributes in stage 1. Ninety–six participants (62 females and 34 males) then rated the attributes in stage 2. All participants were allocated to conditions at random regardless of their sex. Those rating attributes were paid £5 for their participation.

Materials. Eight pairs of social stereotype categories were generated. For each pair, one of the categories was a typically male stereotype, while the other was a typically female stereotype. The pairing thus led to unusual combinations of concepts. The categories are shown in Table 6.

Procedure. As in Experiment 1, in the first phase groups of participants were asked to generate attributes which might be used to describe each of the 16 stereotypes and their 8 conjunctions, while adopting the point of view of either one or the other constituent stereotype. Instructions were as follows:

This is a study about the views people have of each other. You are given a type of person to describe and a point of view from which to describe them. You are asked to give a list of characteristics which you would expect someone with the given point of view to say about the person to be described. There are five spaces for your responses, but please do not feel that you must give five responses. If more than five responses occur to you then add them in the extra space given under each question.

A worked example was given for a category not used in the experiment proper. In the second phase, all attributes generated by any of the groups for a particular pair of categories were listed in alphabetic order, and different groups of participants made judgements about the appropriateness of the descriptions for each of the stereotypes and their conjunctions, again from one or the other point of view.

Design. The design followed the same principles as Experiment 1, with the exception that an additional factor of Gender (of the target not the participant) was added. Thus half the participants received male points of view and male target individuals while the other half received female points of view and female targets. For each group of participants, gender was held constant across point of view and category, and the point of view adopted was always the same gender as the target category being considered. Thus for a combination like Boxer/Nurse, when the target person was male ("a man who is a boxer and a nurse"), the gender consistent point of view was a male boxer, and the gender inconsistent point of view was a male nurse, while when the target person was female the two points of view were those of a female nurse or a female boxer. Thus male points of view of female targets or vice versa were never elicited. The full design for each phase of the experiment involved twelve different sets of judgments, obtained from twelve different groups of participants. In phase 1, seven participants were allocated to each group. Taking the Boxer/Nurse pair as an example, the first group was asked to take the point of view of a male boxer (a gender consistent point of view) and generate attributes to describe a male boxer. A second group judged the same target category (male boxer) but took the point of view of a male nurse (a Gender inconsistent point of view). A third and a fourth group judged the alternative target category (a male nurse) from the same two points of view (male boxer or male nurse) respectively.

The fifth and sixth groups judged the combined conjunctive category ("a man who is a boxer and a nurse") from the same two points of view again. Finally groups seven to twelve repeated the whole design with the same categories, but with female points of view about females, in place of male points of view about males. Since there had been little or no effect of the order within conjunctions in Experiment 1 (and such effects are generally small or absent in attribute generation and rating tasks, Hampton, 1987), order was held constant in the design. (Manipulating order in addition to all the other factors would have led to a design with a total of 48 different groups of participants). For each conjunction, the gender consistent category was always placed first in the conjunction, as in "A man who is a boxer and a nurse" or "A woman who is a nurse and a boxer". This order appeared to be most natural pragmatically, where the unusual category assignment was placed at the end of the phrase where novel information is normally highlighted in speech. Order effects may also be expected to be small given that both categories were in the relative clause of the noun phrase (unlike earlier studies which used head noun plus restrictive relative clause constructions such as "A sport that is a game"). Groups 1–4 and 7–10 generated attributes to all 16 individual categories. Groups 5, 6, 11 and 12 generated attributes for the 8 conjunctions. Phase 2 (attribute rating) followed the same design. Because of missing data and incomplete booklets, additional participants were recruited to bring the N for all cells in the design to a minimum of 8. Each of the 8 (category pairs) x 2 (gender) x 2 (consistency of point of view) x 3 (own, other and conjunction) scales were assessed for reliability. All except for 4 had alpha greater than 0.7. Mean reliability was higher for constituents (.91) than for conjunctions (.78).

Results and Discussion

Mean ratings. The appropriateness ratings obtained in Experiment 2 were analysed to generate a full set of 12 mean ratings for each category pair. These 12 means corresponded to four for each constituent, and four for their conjunction, the four in each case corresponding to the two points of view which could be either gender consistent or gender inconsistent, depending on the gender of the target person.

Gender	GENDER CONSISTENT POINTS OF VIEW		BETA WEIGHTS		R
	Category A	Category B	Own	Other	
Male	Car Mechanic	Reads Romances	.018	.821*	.811
Male	Tractor Driver	Ballet Dancer	–.165*	.722*	.790
Male	Fighter Pilot	Child Minder	.709*	.720*	.838
Male	Road Digger	Does Embroidery	.115	.925*	.852
Male	Football Hooligan	House husband	.042	.649*	.614
Male	Rugby player	Knits	–.036	.665*	.695
Male	Refuse collector	Makes cakes	.090	.703*	.662
Male	Boxer	Nurse	.212*	.701*	.689

Gender	GENDER CONSISTENT POINTS OF VIEW		BETA WEIGHTS		R
	Category A	Category B	Own	Other	
Female	Ballet Dancer	Tractor Driver	.851*	.353*	.783
Female	Child Minder	Fighter Pilot	.531*	.643*	.864
Female	Does Embroidery	Road Digger	.860*	.967*	.678
Female	House wife	Football Hooligan	.232*	1.036*	.874
Female	Knits	Rugby player	.402*	.942*	.876
Female	Makes cakes	Refuse collector	.609*	.965*	.758
Female	Nurse	Boxer	.755*	.542*	.876

Table 6. Regression statistics for predicting attribute ratings for the conjunction from attribute ratings from each constituent, for gender–consistent points of view (Category A as point of view). *Note*: Own = rating for the gender consistent constituent Category A, which was adopted as point of view. Other = rating for Category B which was gender inconsistent.

Thus for the pair Boxer Nurse, the four points of view would be a male boxer or a male nurse (rating male targets) or a female boxer or a female nurse (rating female targets). As in Experiment 1, regression statistics were calculated to investigate the degree to which an attribute's appropriateness for a conjunction was related to its appropriateness for the two constituents of the conjunction. Recall that in the antagonistic pattern shown in Experiment 1, there tended to be a positive regression weight for the degree to which an attribute was judged true of the Other constituent and low or zero weight for the degree to which it was true of one's Own constituent category. Results for Experiment 2 are shown in Table 6 (for Gender–consistent points of view) and Tab;e 7 (for Gender–inconsistent points of view).

Gender	GENDER CONSISTENT POINTS OF VIEW		BETA WEIGHTS		R
	Category A	Category B	Own	Other	
Male	Car Mechanic	Reads Romances	.742*	.317*	.724
Male	Tractor Driver	Ballet Dancer	.507*	.411*	.556
Male	Fighter Pilot	Child Minder	.454*	.664*	.910
Male	Road Digger	Does Embroidery	.637*	.597*	.740
Male	Football Hooligan	House husband	.321*	1.030*	.845
Male	Rugby player	Knits	.561*	.753*	.801
Male	Refuse collector	Makes cakes	.279*	.550*	.504

Gender	GENDER CONSISTENT POINTS OF VIEW		BETA WEIGHTS		R
	Category A	Category B	Own	Other	
Female	Reads Romances	Car Mechanic	.941*	1.160*	.820
Female	Ballet Dancer	Tractor Driver	.280*	.849*	.882
Female	Child Minder	Fighter Pilot	.831*	.495*	.865
Female	Does Embroidery	Road Digger	.570*	.871*	.642
Female	House wife	Football Hooligan	.863*	.587*	.705
Female	Knits	Rugby player	.954*	.723*	.753
Female	Makes cakes	Refuse collector	.534*	.610	.632
Female	Nurse	Boxer	.212*	.731*	.842

Table 7. Regression statistics for prediction attribute ratings for the conjunction from attribute ratings from each constituent for gender inconsistent points of view.

The beta weights were submitted to ANOVA with the 8 category pairs as random variable, and with Gender, Consistency, and Own–Other constituent as three repeated measures factors. There were significant main effects of Gender and of Own–Other, which were included in a significant two–way interaction of Gender and Own–Other ($F(1,7) = 36.95$, $p < .001$). There was also a significant two–way interaction of Consistency with Own–Other ($F(1,7) = 7.13$, $p < .05$). No other effects reached significance. The data are presented in Figure 1.

It is very clear that for three of the four points of view, each constituent plays a strong and positive role in predicting the inheritance of an attribute in the conjunction, as would be predicted by the composite prototype model (Hampton, 1987, 1988). It is only in the case of Male Gender Consistent points of view that beta weights fall. When imagining themselves in a stereotypical male role (e.g. a male boxer), our participants judged that a male who combined male and female stereotypical roles (e.g. a male who was a boxer and a nurse) would not inherit the attributes of the boxer, but only those of the nurse. Although the 3–way interaction did not reach significance ($F(1,7) = 1.26$), it is probable that with a more powerful design the effect would be seen. According to the significant two–way interactions, Male points of view, and Gender consistent points of view both generated greater imbalance away from one's own category towards the other, in the antagonistic pattern of concept combination.

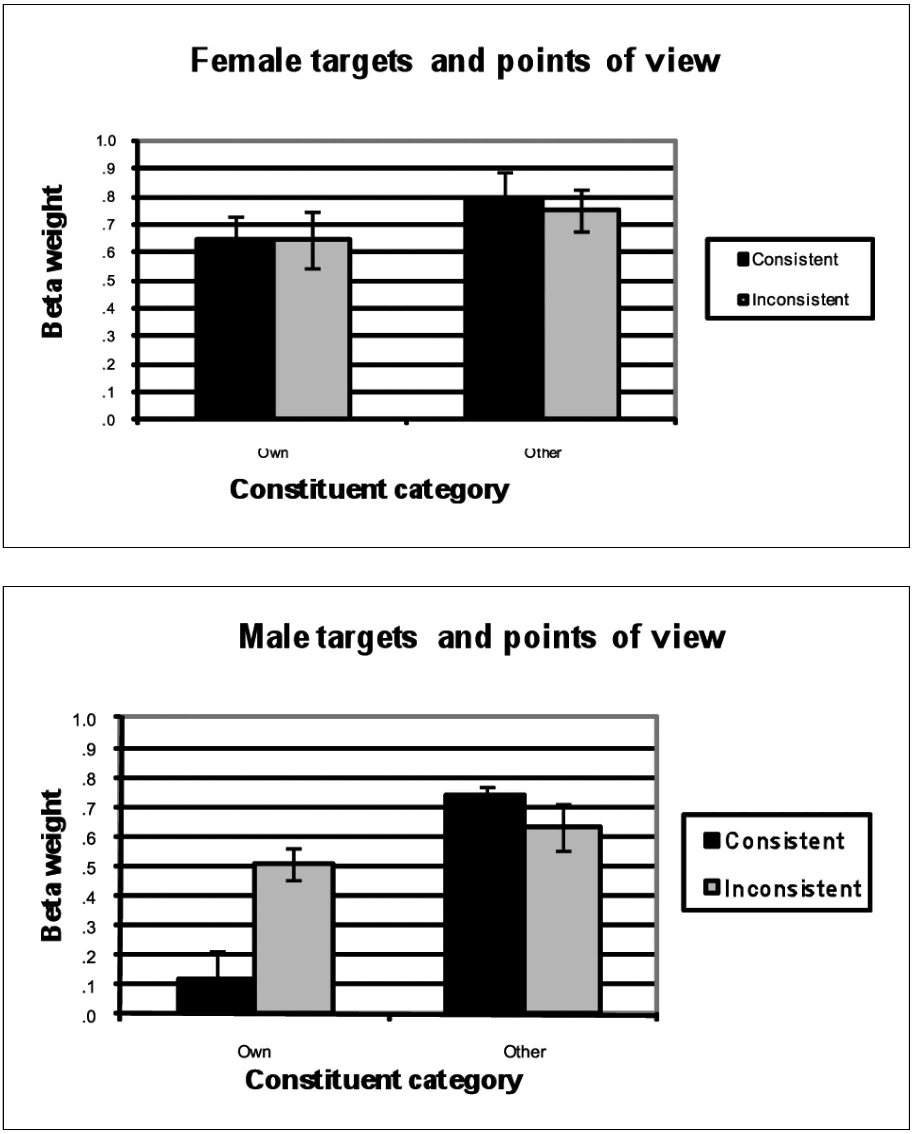

Figure2. Mean beta weights for regressions predicting conjunctive from constituent attribute ratings, for gender–consistent (black bars) and gender–inconsistent (grey bars) points of view.

Attribute inheritance. An analysis was also done of attribute inheritance, similar tothat reported for Experiment 1. To save space it will just be summarized here. Attributes for males considered from gender–consistent male points of view were only inherited 45% of the time from their own category, but 93% from the other category. The balance was more even for gender–inconsistent male points of view (74% for own and 68% for other), and for both female points of view (all between 64% and 76%).

Emergent attributes. As in Experiment 1, emergent attributes were defined as those with positive ratings for the conjunction, that had negative ratings for each constituent. These attributes are listed in Table 8.

Content analysis of these attributes showed that it was much more common for the male points of view to be antagonistic than for the female points of view. Female points of view tended to be more integrational — taking good and bad points from each constituent. The female points of view about females tended to involve many more positively valued emergent attributes than the equivalent male categories (some 34 out of 37 attributes were positive, compared to 8 out of 27 for males). Male points of view appear to be much more prone to take antagonistic attitudes to combining gender inconsistent categories.

CATEGORIES	POINTS OF VIEW			
	MALE		FEMALE	
	CONSISTENT	INCONSISTENT	CONSISTENT	INCONSISTENT
CAR MECHANIC who	Dissatisfied	Dissatisfied	Ambitious	Easygoing
READS ROMANTIC	Elusive	Reliable	Broad–minded	Calm
FICTION		Lonely	Clever	Charming
		Soppy		Caring
				Intelligent
TRACTOR DRIVER	Passive	Confused	Bold	Soft
who is a	Unserious	Dirty	Unconventional	
BALLET DANCER		Eccentric	Enjoys herself	
		Peculiar	Satisfied	
FIGHTER PILOT who	—	Untroubled	Contradictory	Fun loving
is a CHILD MINDER				
ROAD DIGGER who	Single	Unusual	Adventurous	Unstereotypical
does	Multi–talented		Challenger	Unusual
EMBROIDERY	Dare to be different		Just does a job	Positive
	Easy		Healthy	

CATEGORIES	POINTS OF VIEW			
	MALE		FEMALE	
	CONSISTENT	INCONSISTENT	CONSISTENT	INCONSISTENT
FOOTBALL	Homosexual	Vain	Eager	Antisocial
HOOLIGAN	Changeable	Changeable	Sporty	
who is a HOUSE			Unfulfilled	
HUSBAND/ WIFE			Adventurous	
			Football supporter	
RUGBY PLAYER who	—	Brave	—	Organised
KNITS		Funny		Well–rounded
		Eccentric		
		Strange		
		Uncaring of image		
REFUSE COLLECTOR	Confused	Miserable	Simple	Adept
who MAKES CAKES	Articulate	Unsociable	Broad–minded	Clever
	Simple	Unusual	Determined	Enjoys life
		Strange	Equality	Fulfilled
				Multi–talented
				Positive
				Same beneath
				Broad–minded
				Determined
				Fun loving
BOXER who is a	Repressed	Dissatisfied	—	—
NURSE	Uncompetitive	Rival		
	Emotionally split			
	Strange			

Table 8. Emergent attributes with Gender consistent and inconsistent stereotypes.

General Discussion

It would be interesting to consider the meaning of our results for British culture and gender stereotypes. At first view, male points of view appear to be more antagonistic in combining inconsistent social categories. Males were considered less able to belong to gender–incongruent categories without being seen in a negative way by those in only one of the categories. It should however be born in mind that the data here are based on the opinions of people *adopting points of view* rather than the opinions of people actually themselves in those social categories. It is remarkable how people are able to imagine themselves in the position of a socially stereotyped group, and how they can consistently make judgements from that point of view. This skill deserves much more attention as a part of our ability to understand the minds of others. However it would also be valuable to follow up this study with studies of actual members of stereotypic groups (boxers and nurses for example), of each gender.

The primary aim of the study was not to investigate sex stereotyping but to learn more about the processes involved in combining concepts that do not normally combine. In line with previous research (Hutter & Crisp, 2005, Kunda et al., 1990) we have found that interactions occur when unusual or unfamiliar combinations of social categories are formed. More particularly we were able to identify two very different patterns of attribute inheritance. On the one hand, social categories can be combined in an integrative fashion, taking the positive and negative attributes of each category and combining them into a novel composite prototype (see Anderson, 1965). This process is the one that was identified for non–social categories in the earlier work by Hampton (1987, 1997). In order to identify a set intersection of instances in the world, it is necessary to create a set union of the criteria that identify them. Thus, even with vague, prototype–based concepts, the correspondence between intension and extension operates in broadly the same way that it does in classical logic. (To be in the instance class A^B, an item must possess all the necessary features that might be found EITHER in the definition of A, OR in the definition of B).

On the other hand, we have discovered that in certain circumstances, people will resist this integration. Particularly when people imagine the attitude that might be taken by a stereotypical male (e.g. a male Rugby player), they suppose that someone in a sexually ambivalent conjunction that includes both male and female characteristics will not inherit the normal typical properties of the male stereotype. Instead an antagonistic pattern of inheritance appears. The fact that the effect is more strong for male than for female stereotypes probably reflects real asymmetries in gender roles in British society. It is arguably far less easy for a man to adopt female styles of dress or hairstyle without attracting negative attention than for a female to do the equivalent. In British society, it is not at all unusual for females to have short hair, wear baseball caps and jeans, to drink pints of beer and get drunk in the street. Men who might want to wear skirts and high heels and makeup would have a much harder time being accepted.

Experiment 1 confirmed that this is not just a pattern seen in gender stereotyping, but is generally true of social categories that have an antagonistic relationship. Socialist stockbrokers are reviled by socialists as being just stockbrokers, and by stockbrokers as being just like socialists. Neither side is imagined to be willing to acknowledge that the maverick individual would have the positively valued attributes of the group to which they themselves belong. This anatagonistic pattern can be related to the need for conjunctive categories to show a degree of coherence (Kunda et al. 1993). Participants see the irresoluble problems of someone who would be both a socialist and a stockbroker, and so tend to see the conjunction as primarily one or the other (rather like the instability of the Necker Cube). The current studies have shown that this instability can be strongly influenced by the taking of one or the other categories as a point of view.

References

Anderson, N. H. (1965). Averaging versus adding as a stimulus–combination rule in impression formation. *Journal of Exerimental Psychology, 70*, 394–400.

Barsalou, L. W. & Sewell, D. R. (1984). *Constructing representations of categories from different points of view.* (Rep. No. 2). Emory University, Atlanta GA: Emory Cognition Project.

Cohen, B. & Murphy, G. L. (1984). Models of concepts. *Cognitive Science, 8*, 27–58.

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

Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition, 15*, 55–71.

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

Hastie, R., Schroeder, C., & Weber, R. (1990). Creating complex social conjunction categories from simple categories. *Bulletin of the Psychonomic Society, 28*, 242–247.

Hutter, R. R. C. & Crisp, R. J. (2005). The composition of category conjunctions. *Personality and Social Psychology Bulletin, 31*, 647–657.

Kunda, Z., Miller, D. T., & Clare, T. (1990). Combining social concepts: the role of causal reasoning. *Cognitive Science, 14*, 551–578.

Kunda, Z. & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel–constraint–satisfaction theory. *Psychological Review, 103*, 284–308.

McCloskey, M. & Glucksberg, S. (1978). Natural categories: Well–defined or fuzzy sets? *Memory & Cognition, 6*, 462–472.

Murphy, G. L. (1988). Comprehending Complex Concepts. *Cognitive Science, 12*, 529–562.

Murphy, G. L. & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289–316.

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

Osherson, D. N. & Smith, E. E. (1982). Gradedness and conceptual conjunction. *Cognition, 12*, 299–318.

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

Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. (1988). Combining Prototypes: A Selective Modification Model. *Cognitive Science, 12*, 485–527.

Zadeh, L. (1965). Fuzzy sets. *Inform ation and control, 8*, 338–353.

James A. Hampton
Margaret Dillane
Laura Oren
Louise Worgan
City University London
j.a.hampton@city.ac.uk

The Hidden Strengths of Weak Theories

Frank Keil

Abstract There has been a strong tradition of assuming that concepts, and their patterns of formation might be best understood in terms of how they are embedded in theory–like sets of beliefs. Although such views of concepts as embedded in theories have been criticized on five distinct grounds, there are reasonable responses to each of these usual objections. There is, however, a newly emerging concern that is much more challenging to address — people's intuitive theories seem to be remarkably impoverished. In fact, they are so impoverished it is difficult to see how they could provide the necessary structure to explain differences between concepts and how they might form in development. One response to this recent challenge is to abandon all views of concept structure as being related to people's intuitive theories and see concepts as essentially structure–free atoms. The alternative proposed here argues that our very weak theories might in fact do a great deal of work in explaining how we form concepts and are able to use them to successfully refer.

Key words Concepts, intuitive theories, conceptual change, categorization, word meanings, cognitive development

For many years it has been assumed that concepts are embedded within larger systems of beliefs that help articulate their structure. These systems of beliefs are often thought of as intuitive or naïve theories and are thought to be a key way of explaining concept formation and conceptual change. In particular, the emergence of new theories out of old ones in which inconsistencies become apparent are thought to be a primary vehicle for the formation of new concepts (Carey, 2009). In the philosophy of science, it has long been held that theories provide critical frameworks within which concepts are articulated, frameworks that give sense of ontological kinds and of relations between concepts (Kuhn, 1977). In developmental psychology, intuitive theories have even been attributed to infants and have been argued to be the best ways to understand early concepts (Gopnik & Meltzoff; 1997). Intuitive theories have also been seen as causing a potential conflict with more associative views of concepts in which a concept is little more than tabulations of how often features occur and co–occur for certain entities (Johnson & Keil, 2000; Keil, 1989).

Yet, this view of concept formation faces a major challenge. Are the intuitive theories of lay people and of the folk sciences, namely the ways lay people make sense of various phenomena in the world, adequate as means for understanding concept formation, growth and use? The answer is unclear and has led some to propose either minimalist views of concepts (Fodor, 1998) or even to do away with concepts altogether (Machery, 2009). Here I want to suggest that there may be ways to maintain an account in which concepts are associated with rich structures but which also acknowledge the many limitations of intuitive theories.

Let us therefore consider in some detail the view that concepts are embedded in theories, and that they derive their structures from theories. More precisely, concepts are embedded in theory–like structures and are distinguished from each other by the particular ways that each concept is embedded in a web of relations that make up a theory. That web might be characterized as a "web of belief", (Quine & Ullian, 1978), or perhaps more primitively as less belief–like cognitively implicit links to other concepts and properties. Thus, one might argue that more brute force associations between networks of concepts define new concepts (Rogers & Mclelland, 2004). The same idea can be advanced for constituents of concepts, namely that a concept such as DOG is made up of various "perceptual" and "conceptual" features that are presumed to make up the meaning of dog, such as having four legs, barking, being a living creature and having an essence. (There is much to worry about in such accounts, including whether any appeals to features of concepts are in fact simply making more links to other concepts (Fodor,1998), but since much bigger questions will emerge about the larger enterprise, those worries do not need to be dwelt on here.)

One reason for thinking that concepts are made up of theories in this way is because of apparently powerful links between conceptual change and theory change. Whether it be in the history of ideas or a particular child, concepts seemed to travel in groups (Carey, 1985; Keil, 1989; Thagard, 1992). When a child has a particular kinship term that shifts in meaning over the course of development, many others concepts seem to shift at the same time in terms of what they mean to the child (Keil, 1989). When a child's concepts of weight seem to change, concepts of density may change as well, and similarly for heat and temperature (Smith, Carey & Wiser, 1985; Wiser & Carey, 1983). When the concept of evolution by natural selection emerged in the 19th century, it seemed to be related to a change in the concept of a species. Many other examples exist in the history of science and in cognitive development (Thagard, 1992).

In addition to cases of conceptual change, many concepts seem to be interdependent. It is not clear that it is even coherent for someone to claim to have a concept of a mechanical NUT without having the concept of a BOLT that accepts that nut. Similarly. BUY can't seem to stand on its own without the concept SELL, MOTHER without CHILD, and so on. These cases appear to illustrate the idea that concepts are part of larger relational complexes that both give meaning to them and make up their meaning. Come to understand one concept and, in at least some cases, you will automatically understand the other. If concepts are embedded in theories and the same theory applies to two different concepts, it offers an explanation of why they should be linked in understanding and in conceptual change.

The concepts–in–theories idea also seems to gain support from the ways in which notions of "how" and "why" seem to influence all aspects of concept acquisition and use. Often, one of the most striking aspects to concepts is not how often features occur with instances — that is what makes up prototypes (or syndromes, whatever you want to call them), but rather the causal explanatory roles filled by the features associated with concepts. Even very young children and infants do not judge category membership by merely weighing features on the basic of their typicality and then doing some aggregation of such weights over features to determine category membership. The degree to which they think a property is causally important will often trump typicality. Thus, even if all known tires have been black and only 95% have been fully round (as opposed to flat), shape is considered much more important to being a tire than color (Keil, 1994; Keil et al., 1998; Keil, 2010). Similarly, as adults at least, we discount highly salient and reliable features of hair length and clothing as central to a concept of male and female and emphasize other features that are much less frequently observed. In the same way, when we make inductions about other things that are likely to be true of thing, we make those inductions not merely on the basis of past frequencies of features, but also on the basis of guesses about their causal roles in category. So, if told that a sampling of ten cats revealed that all had a particular enzyme for digesting meat and that all had two syllable names, we are much more likely to infer that the enzyme is a critical feature of all cats not the two syllables (Heit, 2000; Proffitt, Coley and Medin, 2000; Wisniewski and Medin, 1994). In both adults and young children causal explanatory knowledge seems to influence how features are used to structure categories and their associated concepts (Hayes & Thompson, 2007; Rehder & Kim, 2006).

In short, concepts as theories (or at least as embedded in theories) seem like a compelling way to characterize their nature in adults. At the same time, there are some concerns in the adult psychological literature as well. For example, frequency based information can have a strong influence in at least some contexts (Hampton, 2000). Moreover, the importance of causal information may vary as a function of domain with it possibly being more influential for living kinds than for artifacts (Hampton, Storms, Simmons & Heussen, 2009). Thus, there are theory–like effects, but they can vary in strength, raising potential questions about how central they are to concepts in general.

Theories and Development

The alleged centrality of theories to our concepts also seemed to be further reinforced by their role in development. Indeed, research in cognitive development was a primary impetus towards this view of concepts. Concepts appeared to change in the course of development in ways that reflected growing webs of belief. Concepts weren't nodes in this network, they were clusters of nodes and links. There was a tacit, but not well articulated assumption, that these clusters were somehow bounded, like one of circles in Figure 1.

As the network of beliefs grew, the clusters of concepts changed, and new concepts emerged. Thus, web growth could spawn new concepts (new clusters) elaborate on old ones in quantitative ways (more links with the same sort of structure) or become more dramatically restructure in qualitative ways (new kinds of structures). There were nagging details about how to determine clusters, how to decide when a new concept emerged, and so on, but these seemed like details that could be worked out.

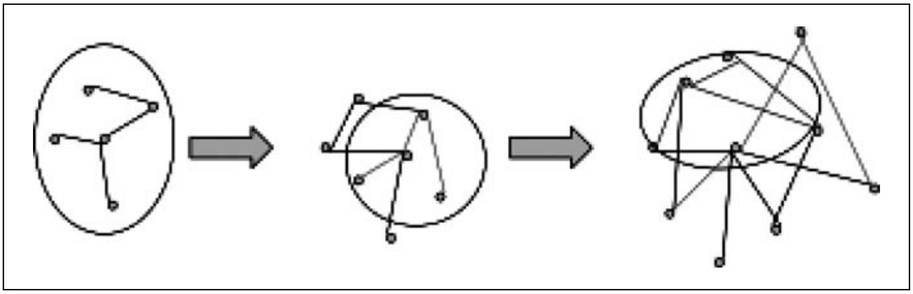

Figure 1. Are concepts regions in networks of beliefs (shown here by ovals), with conceptual change occurring as those networks expand?

Five classes of problems

Other problems, however, were deeper and were more problematic; although here too, many researchers in the field felt that reasonable solutions to these would emerge with time. Most of these were also documented by Fodor (1998, 2005): Fodor & Lepore, 2002). Five classes of problems are particularly relevant.

First, there is "the lost in thought" problem. If concepts were really the same as theories of how all their components worked together, wouldn't theories be too slow? Wouldn't we get lost working through all these theoretical implications? Yet we use concepts quickly and effortlessly. How could our apparent speed of use of concepts be reconciled with the richly textured theories that had to be considered to use them appropriately?

A second problem was that there seemed to be too much change in the theories surrounding concepts, while the concepts seemed to stay the change or only change more modestly. When William Blake wrote "Tiger, Tiger" near the beginning of the 19^{th} century, the state of biology was radically different from what it is today. Evolution was not yet on the horizon, molecular biology didn't exist, and there were many popular misconceptions about the dispositions of tigers. Yet, it isn't unreasonable to say that William Blake and Siegfried and Roy really meant pretty much the same thing when they referred to tigers. How could their tiger concepts be so similar if biological theories had changed so much?

A third problem was that of meaning holism. Where does one theory stop and another begin? As one pursues full and exhaustive explanations of just about any phenomenon one runs the risk of traversing the full extent of the web of beliefs to track down an additional explanatory insight. An exhaustive theory explaining everything about cars might pull on physical mechanics and even quantum phenomena, on chemistry and thermodynamics, on electricity and magnetism, on human physical and cognitive ergonomics (and from there to all of cognitive science and biology), on the economics and geopolitics of fuels, and so on. It would seem we might be at the mercy of a "cognitive butterfly effect." Change a belief at some far off point in the web of belief and why couldn't it traverse back and cause a shift in our core concepts of cars?

There is also the problem of conceptual combinations. If concepts can be understood in terms of the structure of the theories in which they are embedded, shouldn't it be possible to use those structures to predict what happens when they are merged in conceptual combinations, such as pet fish, junkyard dog, and the like? At first glance, it appears that we are not well equipped to explain these combinations. Properties seem to emerge in ways that are not derivable from their alleged internal structures, whether those structures are understood as prototypes or as theories (Fodor and Lepore, 1996; 2002; see also Jönsson & Hampton, 2008). Moreover, it was not clear what it means to combine concepts when those concepts themselves are understood as parts of larger networks of beliefs. Are those beliefs modified or is the circle of interest simply expanded to include both of the constituents with the larger circle of beliefs equal to the new combination? (Neither of these seems very feasible). Moreover, to the extent that meaning holism is a problem, it is an even larger problem in knowing how to combine such massively extended networks.

Finally, there are difficult questions concerning how new nodes in the network of explanatory beliefs might emerge. What does it mean for a network to grow and how could one distinguish between the addition of new beliefs linking up familiar nodes and the actual emergence of new nodes? More concretely, when a child learns the concept of an Apple I–pad, is that child simply welding together a new constellation of beliefs about familiar concepts of electronic devices and music, or does a new node emerge which then can be used as another connector for beliefs? There appeared to be little consensus on how to implement such ideas.

Some Possible Responses to the Five Problems

These problems are hardly new (e.g., Fodor, 1998). For many researchers in cognitive science, however, these did not seem to be insurmountable obstacles to a view of concepts as based on theories. It is far beyond the scope of this paper to go through in any detail the possible ways to address these problems, but a brief mention of some potential ways out is useful because it helps to keep them in mind in considering what may a be a much more profound problem with the concepts–in–theories view.

The "lost in thought" problem might be solved by a mechanism of "pre–compiling". Thus, people might not have to work out all the theoretical implications of a concept in real time and instead could build up a set of useful expectations based on their theories that could then be stored and used quickly in various situations. For example, I do not need to go through the full set of causal explanatory beliefs relating to birds every time I encounter a bird. Instead, over a longer period of time preceding any encounter with a bird, I may have for theory–based reasons, learned to weigh some features such as kind of wings, more importantly than others, such as whether a bird is facing to the right or the left when I first encounter it. Theories could guide one's attention towards certain features over others and in doing so shift even what features are most typically noticed in conjunction with a category. In fact, through such methods as pre–compiling, adults at least can sometimes use theory–based information about features faster than mere frequency based information (Luhman, Ahn, & Palmeri, 2006).

The problem of theory change without concept change might be addressed by discounting the magnitude of theory change for most people. To be sure, the biological sciences have advanced massively in the past 100 years, but perhaps the layperson's concepts of the biological world have changed to a far smaller extent and thus do not pose as much of a problem. If intuitive theories are ones in which concepts are normally embedded, there may be more continuity than change in their nature. There may also be less theory change in the folk sciences where, as we will see, the adult end states are often quite sparse. If the theories are sufficiently sparse, not as much has to change.

Meaning holism might be addressed by a kind of pragmatic pruning that gives locality in real time. Thus, while it is true that explanatory inferences can eventually branch out from a concept to almost anywhere in network of beliefs, perhaps we use some simple heuristics to keep holism in check, such as attaching an exponentially higher cost to traversing each extra link away from the concept we are trying to understand. Moreover, there has been a surge of interest in the philosophy of science to document somewhat similar heuristics used to localize phenomena and rule out some sets of causal relations through simplifying assumptions and idealizations (Elga, 2007; Godfrey–Smith, 2009; Strevens, 2008; Weisberg, 2007).

Conceptual combinations remain a tricky problem, but there are attempts to explain how coarser patterns in domains might actually be being brought into alignment in ways that do enable predictions (Johnson and Keil, 2000; Hampton, 1994). Also, people may have underestimated the extent to which prototypes can foster comprehension of many conceptual combinations (Jönsson & Hampton, 2008). Finally, there are neuropsychological suggestions of how constituents might be superimposed in systematic and predictable ways (Baron, Thompson–Schill, Weber & Osherson, 2010) as well as suggestions of ways in which perceptual simulations might be involved for at least some conceptual combinations (Wu & Barsalou, 2009).

Finally, we might be able to distinguish between core concepts that exist as primary nodes in a network and which exist from early infancy and other concepts that only exist in a derived form based on networks of beliefs to those core concepts. This might provide us with a principled way of deciding how new nodes emerge in the network of beliefs. Thus, there are arguments that core concepts have a special status and character in early development that enable researchers to distinguish them from later concepts that are formed out of these primitives (Carey, 2009; Spelke, 2000; Spelke & Kinzler, 2007).

There remain many major problems to be solved with each of these responses, but they do allow one to believe that the theory–based way of studying concepts could still survive. It seemed reasonable to develop these responses further given the many appeals of the concepts–in–theories view. A much bigger, problem, however emerged when attention turned to the actual complexity of naïve theories.

A Bigger Problem–Weak Theories

From my own perspective, the most difficult problem with the theory–based view of concepts emerged gradually through a series of studies that more frontally tried to ask about what intuitive theories really were like in the minds of others. We began to uncover truly devastating gaps in people's knowledge — gaps that much of the time existed without people have any awareness of them. We called this lack of awareness an "illusion of explanatory depth" (Rozenblit & Keil, 2002). People tend to overestimate how well we can explain things, and children do so to an even more extreme degree (Mills & Keil, 2004). To show this, an experimenter simply asks participants how well they think they know how something works, and then subsequently ask them to explain it (with appropriate training on scales and other experimental design particulars). People are often shocked at how much worse their explanations are than from what they thought they knew. Interestingly, people tend to be much better at assessing their knowledge of how well they know facts, procedures (such as how to make international phone calls), and narratives (how well they know the particular plots of books or movies). In contrast, they are very poor at estimating their knowledge of how and why. The effect has also been found for estimates of understanding of political candidate's explanations (Alter, Oppenheimer & Zemla, 2009).

The failure to recognize the shallowness of our explanatory understandings creates a problem. If the theories are much weaker than we think, how much work can they do? We can retreat to talk of framework theories, or core theories, and kindred kinds of notions (e.g., Gopnik & Wellman, 1992l; Wellman & Gelman; 1992;), but these retreats have their own serious problems. If you look at some of the theories that ascribed to young children — they are at best sometimes 3 nodes and 3 links. The very young child's theory of mind has been characterized as roughly: "I have desires that cause me to engage in actions." That is the entire "theory". A little older child might have the following: "I have beliefs. Beliefs cause desires. Desires cause me to engage in actions." Similarly, very early folk biology might be: "I believe in a vital force. That vital force helps me to move; and if there is some force left over, it helps me grow." (Inagaki & Hatano, 2002). If those simple components are all there is to framework theories, they are not going to do a lot for us in terms of articulating the structure of concepts. If there is more to framework theories, it is not clear what those additional details look like.

To make matters worse, we also have a high tolerance of contradictions. As has been shown repeatedly, people can believe rather large chunks of information that are completely contradictory to each other and not realize it until its explicitly pointed out to them (Chin and Brewer, 1993). To flesh out two examples a bit more, many adults will state that they believe that animal kinds have fixed essences yet also state that they believe in gradual evolution through natural selection (Shtulman & Schulz, 2008). Yet, natural selection can only operate on a species that is defined as a distribution of traits rather than having some set of necessary defining features. In the realm of describing human behavior, the same person can state that that human behavior is s result of strict causal determinism and not free will while also later stating that people are morally responsible for their actions (Nahmias, Coates, & Kvaran, 2007).

Perhaps the happy existence of such contradictions in one person's mind can be seen optimistically as a sign that holism cannot really be at work in real minds (if you automatically traversed the full net of beliefs you would be aware of all the contradictions), but it makes all the more difficult any idea that concepts emerge out of a richly articulated and coherent set of theory–like beliefs. It has been repeatedly suggested that people strive towards coherence and use it to structure their beliefs and certainly some preferences for coherence occur (Thagard, 2000), but at the same time, there are clearly other factors that limit the reach of coherence.

One extreme reaction to contradictory beliefs is to say that there is no overall linking structure to our beliefs, that knowledge falls apart into little tiny pieces. DiSessa (1993; Di Sessa, Gillespie, & Esterly, 2004), for example, argues that our knowledge may be nothing more than a collection of phenomenal primitives, or "p–prims". It may, however, not be necessary to abandon all structure. There may be ways we can talk about a more relational structure, but it just can't be like a traditional "theory".

The core problem may be the following: There is, or most of us, no theoretical difference between lions and tigers. I know they are different, I think I know they mean different things, but I cannot provide a theoretical reason that distinguishes them (Fodor, 1998). How then, can concepts be differentiated in terms of the theoretical frameworks within which they are embedded? I may believe that lions and tigers differ for interesting reasons related to theoretical notions in biology, but do not actually know any of those differences. I do have a weak sense of what matters for the difference, namely DNA and the ways a genetic code leads to proteins and other products that in turn give an animal its properties; but I have no idea of what it is about lion DNA that makes it a lion and not a tiger. So, at best, I have some hunches about the kinds of things that make a difference in distinguishing between these two animal kinds; but I cannot provide any details about them whatever.

In other cases, an understanding of what would make a difference may be even weaker. I may know that something tiny inside gold makes it different from silver, but know nothing at all about the nature of that tiny micro–structural component beyond the idea that it somehow causes gold to behave like gold, and silver like silver. This is what we might call "blind faith essentialism" in its purest form.

Do weak theories force us to accept notions of concepts as atoms, with no constitutive structure (see for example, Fodor 1998)? One strong reason to resist that move is the issue mentioned earlier that concepts travel in groups. If my concept of MOTHER changes, so does my concept of CHILD. Concepts can be mutually parasitic off each other for their meanings in ways that seem to defy the idea that they have no internal structure. In addition, there is the critical centrality of how and why. Why it is that features that co–occur equally for instances of concepts are either ignored or attended to because they fit with some notion of how and why? Similarly why are some correlations between features ignored or embraced in ways that reflect intuitions about their causal centrality to a domain? One way out may be to propose that we do track causal and relational structures in the world in a way that is less theory–like than we used to think. This way of tracking causal structure may often not even be belief–like, but might still work in a manner that supports concept acquisition and use and is part of the concepts themselves.

Tracking Causal Structure

What are some of the alternative ways that we do track causal structure? One very simple way involves knowing what kinds of property types are likely to do important causal work in a domain. Very young children, infants, and even some other primates seem to know, for example, that when one is thinking about tools, shape is going to usually matter more than color. In contrast, when one is talking about foodstuff, color usually matters more than shape (Keil, 1994; Keil et. al., 1998; Keil, 2010; Santos et al., 2002). These intuitions about the relative importance of property can be used to construct " causal relevancy profiles " that help constrain though about members of categories.

A different level could be understood as that of causal powers, knowing that certain classes of things have certain dispositions to produce certain effects. For example, preverbal infants (12 month olds) know that intentional agents have the power to create order out of disorder while non–intentional ones do not (Newman, Keil, Kulhmeir & Wynn, 2010). In that set of studies, infants see a pile of disordered blocks, a barrier comes up, comes down again, and reveals the blocks in a neatly ordered array. Infants think that only an intentional agent can bring about such a change, not something non–intentional, like a rolling ball. If the event is reversed and blocks go from an ordered array to a disordered one, the infants understand that both intentional and non–intentional agents have the power to bring that second kind of change about (Newman et. al., 2010).

At yet another level of causal analysis, young children can think about what kinds of functional interpretations make sense with different sorts of kinds. Thus, if preschoolers are invited to ask questions about novel artifacts and novel animals that they have never seen before, they approach them very differently in terms of the kinds of causal regularities they think are at work (Greif, Kemler–Nelson, Keil, & Guiterrez, 2006). For a novel artifact, they are likely to ask what the artifact as a whole is for: "What's that for?" There are far fewer spontaneous questions of this sort about novel animals — they are unlikely to ask what the animal as a whole is for. The children will, however, ask about what parts of an animal are for. "What are their claws for" or "What is this long beak for?" Even if they have no idea what an animal or machine is called and have never seen it before, they seem to have quite sophisticated expectations about the kinds of relational and causal patterns that go with different domains.

Children use these notions of causal patterns to guide their intuitions about the division of cognitive labor. Thus, even if they don't know who knows what, they know there are different kinds of experts out there that they can defer to (Keil, Stein, Webb, Billings & Rozenblit, 2008). This knowledge may be critical to how they set up concepts when they have almost none of the details themselves. It may provide a sense of the kinds of properties and relations that are important and which experts are likely to know about such matters.

Concepts as Chimeras

Perhaps concepts are best understood as chimeras. They are not simply prototypes, they are certainly not definitions, and they are not theories, yet elements of each of these seem to be at work. There may also be a rich causal relational structure that is part of the story. In addition, it may be necessary to incorporate into concepts and their formation the idea of "locking" (Fodor, 1998). People to lock onto objects in ways that often do not have a rich underlying propositional structure serving as support. They seem to use a grab–bag of components to stably refer, ranging from probabilistic tabulations of features associated with categories to evaluations of trustworthiness in a social network of deference in order to ground their use of words. For example, many people may freely use the word "wombat" but have no idea of wombat perceptual or behavioral features. Yet, they arguably lock onto wombats by being reliably plugged into a network of deference and expertise. Even young children are surprisingly sophisticated at linking abstract causal–relational patterns to broad domains, such as social interactions, artifacts, intentional beings, mechanical agents and the like, and they use those to guide categorization, deference and learning. Locking in this way, with all its variations of methods, may be in place from the earliest moments of word learning.

Consider a concrete example of how locking might use such components. One of the animals in Figure 2 is a weasel and the other is a ferret.

Even though one may have no knowledge of any specific features that distinguish these two kinds of animals, one might nonetheless firmly believe that one has both of those concepts. What does it mean to say

Figure 2. Weasels versus ferrets. One of these creatures is a ferret and one is a weasel. We may believe these to be importantly different kinds, but may have no idea of any particular differences between them.

that a person has two distinct concepts of a weasel and of a ferret yet has absolutely no idea what the difference is between them? One answer suggests that such a person thinks he knows who knows. He thinks he knows who the appropriate experts are and how to access them. He may be mistaken, but his beliefs in such experts are enough to convince him that he "has" two separate meanings (see Putnam, 1975).

One part of this knowing who knows may involve the notion of "sustaining mechanism" (Laurence and Margolis, 2000; Margolis & Laurence, 2003). Sustaining mechanisms are mental operations that enable our concepts to lock onto the appropriate classes of entities. That they exist is self-evident — their relation to concepts is more controversial. Laurence and Margolis discuss three kinds of sustaining mechanisms: 1. those that are theoretical and allow you to lock onto objects, 2. those that are based on deference to experts, and 3. those that are based on a syndrome, or something like a prototype. In most cases, the issue may not be which one of these at work. Instead, all three may usually be involved at the same time in instances of locking. Here is why. Our theories are too weak to work on their own. But often, when we decide whether something is a ferret or a weasel, what we are doing is having a crude notion of who the right expert is and using those weak theories to find the right expert realm. We then use that idea of appropriate expert realm to help us defer to others, and we also then use that deference to determine which features of the syndrome to attend to. No one sustaining mechanism may be enough in most real-world cases.

A critical question is whether the sustaining mechanisms are part of the concept itself, as opposed to just being tools that helps us lock. Are sustaining mechanisms like a microscope that helps us lock but which should not be confused with the locked thing? There are reasons to resist that conclusion. It may be that, even for microscopes, part of what it means to have the concept BACTERIA is to know what kind of tool a microscope is. It is not just an abstract tool in the most general sense. One has to know, for example, that it is a way to get information about invisible microscopic structures, that the microscope has some causal efficacy. Moreover, that understanding may be critical to my concept of BACTERIA. The same holds for experts and thinking of experts as sentient "tools" like the microscope. You cannot point any expert at any object and expect to get the right answer. You have to know what kind of expert you are talking about. You have to know that there are different kinds of experts, who have different specializations in different causal regularities.

It turns out that very young children are sensitive to some of the ways expertise works. Even by the age of 3 or so, they start to know that there are different kinds of experts in the world (Lutz & Keil, 2002; Danovitch & Keil, 2004; Keil et. al., 2008). They know that adults are not omniscient and have different zones of cognitive competence. These children must have some mastery of the causal structure of the world to even be able to engage in the practice of deference and the use of expertise. Thus, it can be shown that they are relying on abstract causal schemas to solve the division of cognitive labor problem (Keil et. al., 2008).

Consider how this all might come together in the acquisition of the concept of a CARBURETOR. This truly a speculative account, but it will serve to demonstrate how the idea might work. (It may be an especially interesting case because soon carburetors will no longer exist. They are vanishing due to their replacement by fuel injection systems). We might hear the word "carburetor", and we might then hypothesis–test whether it is an artifact or a natural kind. Here, we might grant that the notion of artifact is innate as well as perhaps the simplest sense of natural kind (not the more complex sense inherent in the philosophy of science). We might then quickly map the word onto the artifact domain; there are lots of perceptual heuristics to tell whether something is an artifact or not (e.g., Levin, Takarae, Miner & Keil, 2001). This locking onto a broad category such as artifact raises the question of whether there is whether there is a notion of differentiating sustaining mechanisms that allow us to go beyond those broad categories. Thus, initially, our locking is so crude that we really can't have different kinds of concepts below a certain level, but we then come to have them as our locking mechanisms become more refined. This is one way in which concepts become formed out of earlier substrates.

In these cases, to have differentiating concepts is to have differentiating sets of sustaining mechanisms. Those increasingly fine–grained sustaining mechanisms may be what allow us to be more and more successful in picking out appropriate categories. Moreover, the mechanisms may not be sharpened just by hypothesis–testing. They may proceed instead by becoming more and more sensitive to the kinds of causal patterns that are associated with different kinds of experts, that is with increasing ability to pick out different kinds of regularities in the world.

In such accounts, concepts might still be considered as autonomous atoms; however, the sustaining mechanisms are so linked to them that they may not be able to be separated from the concepts proper. What then is the role of theories in all of this, especially if theories are so weak? How do weak theories strongly constrain? It may be that weak theories do set up boundary conditions on concepts, namely that if you don't have the abstract causal patterns that tell you the proper domain of a concept, you simply don't have the concept. Someone who thinks that carburetors do have micro–structural essences, that they have no overall function, or are non–physical, simply may not have the concept. Plenty of more detailed beliefs could be wrong, but someone cannot go so far as to violate these overarching patterns. Someone could have mistaken beliefs about the shape, the local function, or the material substrate of carburetors, but they are not licensed to have any mistaken beliefs at all (see also Keil, 1979). A different set of beliefs will be at work for living kinds, such as that they do have micro–structural essences and that they do not have functions as wholes even as their parts can have functions People will therefore have different expectations for living kinds that also cannot be violated. Weak theories do not directly tell lions apart from tigers. But they may provide guidance to deference and ways of access to information. They may guide the construction and the differentiation, of domain–specific sustaining mechanisms and, in this way, are involved in accounts of concept formation.

It is difficult to distinguish this sort of account from Fodor's atomism in which sustaining mechanisms are exterior to the concepts themselves. There are no easy algorithms for telling what is in the concept proper versus what might be in a distinct enabling cognitive structure. But there seem to be powerful constraints in terms of causal–relational patterns that we all pick up on from an early age and which we see as fitting with high–level domains. These are domains like living kinds, artifacts, and intentional agents. These domains are not like traditional theories. Indeed we may sense many of the patterns associated with those domains at a highly implicit level that is only revealed when we look at what information children and adults must be aware of to solve certain tasks. Somehow, we have to learn how to look at equally typical features and weigh them differentially because of beliefs about their causal centrality, and then use that information to guide locking. As a first pass, all of that might still be best thought of as part of the concept proper.

A final issue concerns whether compositionality clearly argues against having sustaining mechanisms being parts of the concepts themselves. It is not at clear, for example, that even the most exhaustive analysis of sustaining mechanisms for two concepts would allow us to explain how they compose. What, for example, is the relation between the sustaining mechanisms for red things and for shirts, that allows us to pick out red shirts? If the mechanisms are parts of the concepts, should not the structure as revealed by such mechanisms enable us to predict the ways in which concepts compose? Given how difficult it is for any approach to provide full accounts of composition, it is not clear that sustaining mechanisms are especially vulnerable. After all, we think that hydrogen and oxygen gas molecules are usefully described in terms of their constituent atoms and bonding relations even though it remains a mystery to fully explain the properties of water from their combination. The inadequacy of some alleged constituents of concepts to fully explain conceptual combination may therefore not be sufficient grounds for dismissing their roles as constituents of concepts. If other phenomena, such as conceptual change and concept formation, can be usefully understood in that matter, then perhaps that is enough.

Conclusion

I have tried to sketch out how the concepts–in–theories view, while very appealing, also has serious limits when one considers the minimalist nature of many intuitive theories. In the end, the theory–like effects associated with so many aspects of concept acquisition and use, argue for still trying to have a way in which causal explanatory structure is part of concepts. It may be that such a route exists through the ways in which weak theory–like structures guide notions of expertise, deference, and feature centrality. It is far too early, however, to know whether this sort of program will provide a fully satisfactory answer to the problem of what sorts of structures make up concepts and can explain their formation in development and learning.

References

Ahn, W. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition* 69, 135–78.

Alter, A. L., Oppenheimer, D. M., & Zemla, J. C. (2009). A construal–based mechanism for the Illusion of Explanatory Depth. Paper presented at the Society of Judgment and Decision Making annual conference, Boston, MA.

Baron, S., G., Thompson–Schill, S. L., Weber, M., & Osherson, D. (2010). An early stage of conceptual combination: Superimposition of constituent concepts in left anterolateral temporal lobe. *Cognitive Neuroscience* 1, 44–51.

Boyd, R. (1999). Homeostasis, species, and higher taxa. In *Species: New Interdisciplinary Studies*, ed. R Wilson, pp. 141–85. Cambridge, MA: MIT Press

Carey, S. (1985). *Conceptual Change in Childhood*. Cambridge, MA: Bradford Books, MIT

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, 63, 1–49.

Danovitch, J., & Keil, F.C. (2004). Should you ask a fisherman or a biologist? Developmental Shifts in Ways of Clustering Knowledge, *Child Development*, 75, 918–931.

di Sessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, 10, 165–255

di Sessa, A., Gillespie, N., & Esterly, J. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science*, 28, 843–900.

Elga, A. (2007). *Isolation and folk physics. Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Huw Price and Richard Corry, eds. Oxford University Press.

Fodor, J. A. (1972). Some reflections on L. S. Vygotsky's thought and language. *Cognition*, 1, 83–95.

Fodor JA. 1975. *The Language of Thought*. NewYork: Thomas Crowell.

Fodor, JA (1998). *Concepts: Where cognitive science went wrong*. Oxford: Clarendon Press.

Fodor 2005, meaning and world order

Fodor, J. A., & Lepore, E. (1996). The Red Herring and the Pet Fish: Why Concepts Still Can't Be Prototypes. *Cognition*, 58, 253–270.

Fodor, J. A., & Lepore, E. (2002). *The compositionality papers*. Oxford: Clarendon Press.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought.* Oxford: Oxford University Press.

Goldman, A. (2002). *Pathways to knowledge*. Oxford: Oxford University Press.

Godfrey–Smith, P. (2009). Models and Fictions in Science," Philosophical Studies 143; 101–116.

Gopnik, A., & Wellman, H. M. (1992). Why the Child's Theory of Mind Really is a Theory. *Mind and Language*, 7, 145–71.

Greif, M., Kemler–Nelson, D., Keil, F.C. and Guiterrez, F. (2006). What do children want to know about animals and artifacts?: Domain–specific requests for information. *Psychological Science*, 17(6), 455–459.

Gopnik, A. & Meltzoff, A. (1997). *Words, thoughts and theories*. Cambridge: MIT Press.

Greif, M. L., Kemler Nelson, D. G., Keil, F. C., & Gutierrez, F. (2006). What do children want to know about animals and artifacts? Domain–specific requests for information. *Psychological Science*, 17, 455–459.

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

Hampton, J.A. (2010). Concepts in Human Adults. In D. Mareschal, P. Quinn and S. E. G. Lea (Eds.). *The Making of Human Concepts*, (pp. 293–311). Oxford: Oxford University Press

Hampton, J.A., Storms, G., Simmons, C.L., & Heussen, D. (2009). Feature Integration in Natural Language Concepts. *Memory & Cognition*, 37, 1721–30.

Harris, P. (2002). What do children learn from testimony? In P. Carruthers, S. Stich & M. Siegal (eds.), *The Cognitive Basis of Science*. Cambridge: Cambridge University Press, 316–334.

Hayes, B. K. & Thompson, S. P (2007).Causal relations and feature similarity in children's inductive reasoning. *Journal of Experimental Psychology: General*. 136, 470–484.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review*, 7, 569–592.

Inagaki, K., & Hatano, G. (2002). *Young children's naive thinking about the biological world*. New York: Psychological Press.

Johnson, C. & Keil, F.C. (2000). Theoretical Centrality vs Typicality in Conceptual Combinations, in Keil, F.C. and Wilson, R.A. (Eds.). *Explanation and Cognition*, (pp. 327–360) Cambridge, MIT Press

Jönsson, M.C. & Hampton, J.A. (2008). On prototypes as defaults. (Comment on Connolly, Fodor, Gleitman and Gleitman, 2007). *Cognition*, 106, 913–923.

Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective*. Cambridge, MA: Harvard University Press.

Keil, F.C. (1989). *Concepts, Kinds and Cognitive Development*. Cambridge, MA: MIT Press

Keil, F.C., (1994). Explanation Based Constraints on the Acquisition of Word Meaning, *Lingua*, 92, 169–196.

Keil, F.C. (2010). The Feasibility of Folk Science. *Cognitive Science*, 34, 826–862.

Keil, F.C, Smith, C.S., Simons, D., and Levin, D. (1998) Two dogmas of conceptual empiricism, *Cognition*, 65 pp. 103–135

Keil, F.C., Stein, C., Webb, L., Billings, V.D., & Rozenblit, L. (2008). Discerning the Division of Cognitive Labor: An Emerging Understanding of How Knowledge is Clustered in Other Minds. *Cognitive Science*, 32, 259–300.

Kuhn, T.S. (1977). *The Essential Tension. Selected Studies in Scientific Tradition and Change*, Chicago: University of Chicago Press.

Laurence, S., & Margolis, E. (2002). Radical Concept Nativism. *Cognition*, 86, 22–55.

Levin, D.T., Takarae, Y., Miner, A., & Keil, F.C. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. *Perception and Psychophysics*, 63, 676–697.

Luhmann, C. C., Ahn, W., & Palmeri, T. J. (2006). Theory–based categorization under speeded conditions. *Memory & Cognition*, 34, 1102–1111.

Lutz, D. J., & Keil, F. C. (2003). Early understandings of the division of cognitive labor. *Child Development*, 73, 1073–1084.

Machery, E. (2009). *Doing without concepts*, Oxford, England: Oxford University Press.

Mandler, J. M. (2004). *The foundations of mind*. Oxford, England: Oxford University Press.

Margolis, E. & Laurence, S. (2003). In S. Stich & T. Warfield (eds.) *The Blackwell Guide to Philosophy of Mind*, pp. 190–213. Blackwell Publishers.

Mills, C.M. & Keil, F.C. (2004). Knowing the limits of one's understanding: The development of an awareness of an illusion of explanatory depth. *Journal of Experimental Child Psychology*, 87, 1–32.

Mills, CM, & Keil, FC (2005). The development of cynicism. *Psychological Science*, 16, 385–390.

Murphy, GL (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Nahmias, E., Coates, J., & Kvaran, T. (2007). Free will, moral responsibility, and mechanism: Experiments on folk intuitions. *Midwest Studies in Philosophy*, 31, 214–242.

Newman, G.E., Keil, F.C., Kuhlmeier, V. & Wynn,K. (2010). Early understandings of the link between agents and order, Proceedings of the National Academy of Sciences, 107, 17140–17145.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category–based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26 (4), 811–828

Putnam, H. (1975). The meaning of "meaning". In K. Gunderson (Ed), *Language, mind, and knowledge* (Vol. 2, pp. 131–193). Minneapolis: University of Minnesota Press.

Quine, W.V.O., & Ullian, J.S. (1978). *The web of belief*. New York, NY: Random House.

Rehder, B., & Kim, S. W. (2006). How causal knowledge affects classification: A generative theory of categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 32, 659–683.

Rogers, T. T. & McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach*. Cambridge, MA: MIT Press.

Rozenblit, L. R., & Keil, F. C. (2002). The missunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, 26, 521–562.

Santos, L.R., Hauser, M.D. & Spelke, E.S. (2002). Domain–specific knowledge in human children and non–human primates: Artifact and food kinds. In *The Cognitive Animal*. (Eds. M. Bekoff, C. Allen & G. Burghardt). Pp. 205–216, Cambridge: MIT Press.

Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science*, 32, 1049–1062

Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition*, 21, 177–237

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, 10, 89–96.

Spelke, E. S. (2000). Core knowledge. *American Psychologist*, 55, 1233–1243.

Strevens, M. (2008). *Depth: An Account of Scientific Explanation*. Harvard: Harvard University Press.

Thagard, P. (1992). *Conceptual revolutions*. Princeton, NJ: Princeton University Press

Thagard, P. (2000). *Coherence in Thought and Action*. Cambridge, MA: MIT Press.

Vosniadou, S. & Brewer, William F. (1987). Theories of knowledge restructuring in development. *Review of Educational Research* 57: 51–67.

Weisberg, M. (2007). Three Kinds of Idealization. *The Journal of Philosophy*, 104 (12) 639–59.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: foundational theories of core domains. *Annual review of psychology*, 43, 337–375.

Wiser, M., & Carey, S. (1983). When heat and temperature were one. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 75–98). New York: Academic Press.

Wisniewski, EJ, & Medin, DL (1994). On the interaction of theory and data in concept learning. *Cognitive Science*, 18, 221–281.

Wu, L.L, & Barsalou, L.W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psych*ologica, 132, 173–189.

Note: Some of the research reported on in this article was supported by NIH grant # R37HD023922 to F. Keil.

Frank C. Keil
Yale University
frank.keil@yale.edu

Concepts as General Representations in Situated Theories

ELISABETTA LALUMERA

Abstract This paper addresses the issue of the role of concepts as representations of general knowledge, which seems intuitive, but it is seldom explained. I identify two kinds of general knowledge, namely, constitutively general (possessed by all members of a category) and behaviourally general (that can be applied to all members of a category). I review the ways in which traditional theories of concepts have coped with generality, then I focus on situated or 'embodied' theories, which present themselves as highly revisionary with respect to other models. I argue that they face no special problem with the generality of concepts, but the reason is that they buy into traditional models for the explanation of generality.

Keywords Concepts, embodied cognition, generality, categorization.

1. Introduction

Traditionally, concepts are general representations of categories, they encode knowledge that can be projected from one thing of a certain kind to all things of the same kind. Metaphorically, concepts are the 'mental glue' that connects our past experience to the present cognitive task. Arguably, in order to perform this function concepts ought to be stable entities, largely unaffected by contextual applications. Philosophers and psychologists have mostly agreed on some version of this requirement, with different emphasis. Recent studies in the psychology of categorization, in neurophysiology and in the structure of cognition, however, point to the claim that conceptual tasks distributed across the sensorimotor system. It is widely claimed (both by proponents and by critics) that this emerging paradigm of situated cognition is highly revisionary and not compatible with most tenets of the traditional view of concepts. On these grounds, given that generality and the mental glue function of concepts were successfully explained by traditional theories, one may wonder whether a situated cognition theory could do as well. This paper focuses on this question, and I will defend an affirmative answer. As the question of generality and the mental glue function of situated concepts have been brought up by Barsalou (1998) first and at some length, and then mainly by Prinz (2002), I will discuss their positions. I will argue that situated cognition theories of concepts are no worse off than their traditional competitors as far as generality and the mental glue function are concerned. This, however, is due to the fact that they are more similar to their traditional competitors than their proponents declare. Though contextual representations are involved in the process of categorization, most of the (conceptual) work is done by stable entities, as in traditional theories. If I am right, then, situated cognition theory of concepts are adequate vis à vis generality, but they are less revolutionary than they purport to be.

The paper will be organized as follows. Section 2 is dedicated to the intuitive but somehow murky issue of generality and the mental glue function, and in section 3 two traditional ways of coping with generality are individuated and described. In section 4 I will concentrate on Barsalou's theory of perceptual symbols, and in section 5 on Prinz's proxytypes. My point is that their proposals are instantiations of traditional models of generality, and stable representations — though of a particular kinds — play a pivot role in their theories too. Being more sketchy, Prinz's proposal is more vulnerable to objections, but I see no reason why it couldn't be integrated with Barsalou's in a more composite model, which would reconcile the mental glue function of concepts with some amount of context–sensitivity.

2. Generality

The idea that concepts are representations of *general* knowledge about categories is very deeply rooted in our use of the term 'concept', both in philosophy and psychology. Here I will take knowledge to be, in a very broad sense, information or data about some kind of individual or kind possessed by some mind[1]. This broad sense is obviously different from the characterization of knowledge in epistemology, which is much more demanding, as it involves — according to most accounts — justification and truth. As for general knowledge, in a first plausible sense it is knowledge about all (or most) members of a category, *versus* knowledge about individual members. We have a concept of, say, dog, when we possess knowledge about the category of dogs rather than *just* knowledge about individual dogs we happened to have met. This characterization, however, is still ambiguous between two different notions. First, general knowledge can be knowledge of properties that all members of a category possess. For example, a definition of square specifies the properties that all squares possess, therefore knowing a definition of square counts as having general knowledge about squares. However, as prototype theorists and exemplar theorists showed, some thirty years ago, for many categories there are no relevant properties possessed by all members. The well–known example, taken from Wittgenstein and exploited by Eleanor Rosch, is the category of games. Rather than sharing a set of common properties — like being funny or being competitive — things that are games bear complex relations ('family resemblances') one to the other. Moreover, many other categories are such that, even if shared properties of all members exist, they definitely fall out of our cognitive reach. This is most plausibly the case for natural kind terms, like gold and tiger[2]. Thus, at least in the great majority of cases, our concepts do not represent general relevant categories[3].

If it is not knowledge of common properties of category instances, what is general knowledge then? In a second sense, it is knowledge that *can be applied* to all category instances. Take for example a prototype of dog — a statistical representation of dog features. It can count as general knowledge about dogs because it can be applied in categorization and in other cognitive tasks to all putative members of the category of dog.

1. According to Jerry Fodor the only information vehiculated by a concept is the causal relation it bears with a (generally extramental) category. Thus, Fodorian concepts encode knowledge at the minimum level (in fact, Fodor's use of the term is intentionally out of line with contemporary cognitive psychology). See Fodor (1998), (2009).

2. References for these points are Wittgenstein (1953), Rosch & Mervis (1975), Kripke (1972), and Putnam (1975).

3. See Rosch & Mervis (1973), and Smith & Medin (1978) for prototypes and exemplars respectively.

On this view, each particular cognitive encounter with a dog would be mapped with that prototype, namely, the same body of statistical data stored in the long–term memory of the subject (or roughly the same, as prototypes can be modified by experience to some degree). Properties of the prototypical dogs need not be shared by all other dogs, but any other dog can be cognitively accessed by mapping its properties to those of the prototypical dog[4]. I have therefore distinguished between two possible meanings of 'general knowledge', namely, *behaviourally general* knowledge of categories — knowledge that can be applied in dealing with all instances — versus *constitutively general knowledge*, that is, knowledge of features that all instances possess.

With a better grasp of what generality may consist in, one may still ask what is the point of characterizing concepts as representations of general knowledge at all. Why generality of concepts should be a desideratum for a theory of concepts? The same question, from a cognitive point of view, would be: What's the point, for a human mind, of having general representations of categories? These are very old questions. Traditionally, universal terms and general terms have always been a conundrum for philosophers. One suggestion made by Locke (somehow in passing) was that we have general representations because 'it is beyond the power of human capacity to frame and retain distinct ideas of all the particular things we meet with' (1690/1964: 14). This 'just can't do otherwise' kind of reply, however, is not really explanatory as it stands. It needs to be supplemented with some hint at what ideas (be them general or particular) are for. That is, we need to focus on the roles that concepts play in our cognitive system. Concepts as representations of general knowledge are in play both when we learn from experience, and when we apply to experience what we have learned before. The two processes are often called 'abstraction' and 'category induction' respectively. Here is how the psychologists Lawrence Barsalou and Paul Bloom describe abstraction:

[abstraction is] the general ability to generalize across category members [...]. All theories agree that people state generics, such as 'cats have fur', and quantifications, such as 'some mammals swim'. Behaviourally, people produce abstractions" (Barsalou 2003: 1177).

4. If knowledge is stored in the format of exemplars, it may be less invariant across contexts, as Barsalou (2003) stresses.

> You drink orange juice, and you like it. You drink oil, and you don't...These events provide valuable lessons...But you can learn from these events only if you have some mental representation of the relevant kinds. To learn from the juice episode, it is not enough to know that this liquid at this time is tasty; you have to be able to generalize to other liquids. A creature without concepts would be unable to learn and would be at a severe disadvantage relative to creatures that did have these sorts of mental representations (Bloom 2002, 147).

And here is how induction is presented by Gregory Murphy, and by Barsalou again:

> If a friend calls me up and asks me to take care of her dog, as she cannot get home, I know pretty much what to expect. Even if I have never met that individual dog, I do know about dogs in general and what care they require. I don't have to ask whether I should water the dog, feed it, vacuum it, cultivate around its roots, launder it, and so on, because I already know what sorts of things dogs need. In fact, it is exactly this sort of inference that makes categories important. Without being able to make sensible inferences about new objects, there would be very little advantage in knowing that something is a dog or couch or tree (Murphy 2002: 243).

> Once something is interpreted as a COMPUTER, inferences follow, such as that it requires electricity, can be used for e–mail, is easily breakable and so forth. If the object were interpreted instead as SOMETHING THAT THIEVES STEAL, different inferences would follow (e.g. The computer should be locked to its table) (Barsalou 2003: 1178).

In very simple terms, abstraction is the bottom–up process from things to the mind, while induction is the top–down process from the mind to things. Abstraction has to do with knowledge storage (learning), while category induction concerns knowledge delivery (expertise). They are key processes of our cognitive system, as they are activated on–line during categorization and language understanding, and off–line during inference, imagination, and problem solving.

It is sometimes said that in performing the abstraction and the induction function concepts are the 'mental glue' that connects our past experience with the present. In Murphy's example above, the concept of dog mediates between the previously stored knowledge about dogs, and the present cognitive task of predicting what the author's friend's dog will need. As Millikan (2000) observes, if cognition *were* inferential, concepts *would* play the role of middle terms of syllogisms — DOGS eat meat, my neighbour's pet is a DOG, my friend's pet eats meat[5]. In this toy example of a cognitive process, knowledge represented by the concept of dog is general enough to be applied to an indefinite number of different cases, and in particular to the present one. That is why concepts need to encode behaviourally general knowledge, namely, in order to perform the mental glue function required for abstraction and category induction.

3. Models of generality

Now let us switch from a purely functional characterization of concepts to a slightly less abstract level of description of how conceptual functions might be implemented. Theories of concepts over the decades have proposed models of how generality can be represented. My aim here will be to briefly illustrate these models, and to see whether situated concepts in Prinz's and in Barsalou's versions instantiate any of them, or constitute a new viable solution, or rather they are not adequate for explaining generality.

Models of generality reduce to two broad categories, namely, invariant symbols, and summary representations. A well–known exemplification of the invariant symbols model is Fodor's language of thought theory. On this view, whenever a human mind interacts with a certain category, say, dogs, a particular symbol–type gets tokened. While tokens physically vary from context to context (just like spoken words slightly vary in phonological form from utterance to utterance), types are invariant across contexts and across different individuals. Thus, one's previous knowledge that dogs eat meat, and one's current obligation to take care of a friend's dog are represented as strings containing the same symbol–type, namely DOG. We may figure them out as a set of sentences in the language of thought, like <DOGS EAT MEAT>, and <I HAVE TO TAKE CARE OF ANNA'S DOG>. The symbol DOG is a general concept, because it occurs whenever information about dogs occurs. It is general in the strong constitutive sense individuated before, because it conveys some property that all dogs have in common, namely, being a dog.

5. The conditional form signals that this is just an analogy. Cognition is not thoroughly inferential, because not all cognitive contents are propositional, i.e., encoded in true/false statements.

It may be not not as simple as it seems. In fact, as Millikan remarks, the activation of identical symbol types is not the same an a (subpersonal) act of recognition that the information thereby vehiculated is about the same category. The two occurrences of the symbol DOG in the example above need to be recognized as tokenings of the same type for the system to perform the mental glue function. Abstraction and induction, on this model, clearly depend upon a recognition of this sort. Assuming some minimal principle of cognitive economy, however, it is reasonable to expect that our conceptual system is capable of performing acts of identification over type-identical representation vehicles (Millikan 2000: 137). On the invariant symbols model, acts of this sort are plausibly described as symbol-matching. Thus, on this hypothesis, abstraction and induction would exploit some second-order fast (and obviously subpersonal) mechanism of symbol-matching.

There is no second-order process involved in the summary representation model. During abstraction, the summary representation is directly matched with the on-line representation, for friend's golden retriever coming at the door. A summary representation is a listing of features that all, most, many or some members of the relevant category possess. The shift from the universal quantifier to the existential corresponds to the shift from the classical theory (definitions) to prototype theories (statistical representations). Protototype theories would deny that a core set of properties is common to all members of a category, but nonetheless a prototype encodes knowledge of properties that all members can possess, or are likely to possess in different degrees. There has been a constant evolution within prototype theories. The seminal idea proposed by Eleanor Rosch (1973) was that summary representations that encode general knowledge of categories were representations of ideal members, namely, members possessing all the properties normally found in the relevant category. Starting from Rosch and Mervis' (1975) updated version of the proposal, however, some of the properties (features) represented in a prototype are more important than others, that is, they are weighted. A property will have a higher weight if it appears very often in a category and does not appear in others — for example, for the category of dogs the property of barking has very high weight, while the property of eating meat has a low weight, because only dogs bark, whereas many other animals eat meat. Introducing weights within summary representations is particularly important for categories whose members vary a lot along different dimensions. For example, tennis rackets do not vary a lot, while dogs do (think of the colour, hair length, size). A summary representation with weights of the category of dogs would have — for example — many possible sizes represented, with high weights for the central area of the scale of sizes (say, the size of cocker spaniels) and lower weights for the top (st. Bernard shepherd), and the bottom (chihuahuas). On this view, the same weighted prototype is common to all dog categorization. This would make it possible to accumulate knowledge from German shepherd examples, and reuse it to perform induction over chihuahuas. More recent versions of summary representation models add structure to weights, that is, encode knowledge about how the different properties that members of the category are likely to possess are related one to the other[6]. In general, the question of how much information about a category can be represented by a summary representation is an open one.

Concepts as invariant symbols in Fodor's theory are pure indicators of a mind's cognitive contact with a category, they do not convey any other information than that. To use another metaphor, they are labels. They are, however, components of strings that represent knowledge. In the invariant symbols model, therefore, the function of concepts — i. e. to represent knowledge of categories to be employed in various tasks — is performed by the whole conceptual system. Knowledge about the category of dogs is the totality of language of thought strings with the symbol DOG in them. Traditional summary representations are more informative than Fodorian invariant symbols. Alternatively, one may characterize invariant symbols as a limit case of summary representations. Both the invariant symbols model and the summary representations model, however, share an important characteristics, namely, the fact that one and the same type of representation (be it atomistic or complex) gets tokened when a certain category is processed. This has always been a mainstream assumption in the study of concepts. As Keil (1994: 169) writes, «shared mental structures are assumed to be constant across repeated categorizations of the same set of instances and different from other categorizations. When I think about the category of dogs, a specific mental representation is assumed to be responsible for that category».

In other words, both invariant symbols and summary representations are 'sameness markers' in Millikan's sense. Something is a sameness

6. For schematas see Markman (1999).

marker if 'any information derived from the same thing in the environment shows up marked by this marker' (of course, when everything goes well) (Millikan 2000: 141). Thus, on Fodor's theory, DOG is the sameness marker for any piece of knowledge pertaining to dogs, whereas in the summary representations model any piece of knowledge pertaining to dogs is matched with the prototype or schema of dog. Being sameness markers, concepts as invariant symbols and concepts as summary representations perform the mental glue function required for abstraction, and for category induction.

4. Perceptual symbols and generality

The basic units of Barsalou's 1998 theory of concepts are called 'perceptual symbols', and they are characterized as follows:

- origin: they derive from the operation of selective attention to perception.. Information from perception is filtered out by selective attention, and then stored in long term memory.
- ontology: they are records of neural activation that arises during perception[7].
- function: they represent schematic components of experience in any modality (visual, auditory, proprioceptive, introspective)[8].

Perceptual symbols are not concepts. On Barsalou's view, the functions of a conceptual system are performed by data structures called 'simulators'. In abstraction processes, the simulator for a category integrates perceptual symbols that are extracted from instances of that category. For example, perceptual symbols derived from perception of a car door, a car window, and a car inside are stored together. In induction, the simulator reproduces some of these perceptual symbols (the process is called 'reenactment') given the specific contextual input. For example, upon hearing the question 'what colour is your car?' the simulator for your

7. More precisely, they are neural correlates of perceptual contents, a neural correlate being defined as the smallest region of the brain cortex the activation of which is sufficient to produce the state in question.

8. Plus, they are generally not conscious.

car activates the perceptual symbol of the car body, while the perceptual symbol of the inside remains inert (usually). Or alternatively, upon hearing a certain noise of your car engine it may activate the introspective feeling of worry that something goes wrong (and every other package of information about cars and engines may remain inert). Thus, the simulator makes stored knowledge available for dealing with the current task involving the category in question, but also makes a selection of what is relevant. Indeed, what is employed in a given cognitive task is a small portion of the knowledge stored in a simulator — it is just as much information as the ongoing process requires. This is the much emphasized contextual character of the theory of perceptual symbols.

As for concepts, Barsalou is not so keen on defining them, because he thinks that the main aim for a psychological theory is to understand the mechanisms involved in categorization, however these may be called (Barsalou, Kyle Simmons, Barbey, and Wilson 2003: 84). The idea is that simulators are equivalent to concepts, whereas simulations or reenactments correspond to what is usually called 'conceptualization' or 'conception' (Barsalou 1998, § 2.4.3, Barsalou et al. 2003: 89[9]).

So far, what we've got is an account of what simulators do, not of how they do that, namely, how they integrate perceptual symbols during abstraction and select them for category induction. How do perceptual symbol systems represent generality? Clearly, simulators are not summary representations. A closer look will show, however, that the other model of generality is at play, namely, the invariant symbols model. Let's see the details.

On Barsalou's view, integration of perceptual symbols is performed by conjunctive neurons. Perceptual symbols that occur across different presentations of the same category are captured by conjunctive units of neurons in modality–specific systems, and correlated feature patterns in different modalities (e.g., visual and auditory) are captured by higher–order conjunctive units in more integrative crossmodal systems. This is how abstraction works. Moreover, particular conjunctive neurons are tuned to the occurrence of particular classes of perceptual symbols, so

9. Barsalou and colleagues often characterize simulators are concept types, and simulations are concept tokens. I prefer not to adopt this terminology because simulations are partial activations of the knowledge stored for a category, not just contextual activations.

that they are able to reactivate those symbols when the category instance or the specific feature is no longer available — this is how induction works. Both in abstraction and in induction, therefore, patterns of activation of the conjunctive neurons tag — so to say — perceptual symbols with the same external origin. Conjunctive neurons are the sameness markers of the theory of perceptual symbols, and they perform the mental glue function both in abstraction and in induction. They are what perceptual symbols from the same category have in common, and they differ from invariant symbols of traditional theories only in that they do not belong to a dedicated system, the conceptual system, as they are part of the perceptual and motor system (but this feature is not relevant as far as their role for generality is concerned) [10].

Of course, the theory's adequacy relies on an empirical issue, namely, the existence of conjuntive neurons in perceptual and motor areas of the brain, which has been advanced by A. R. Damasio and colleagues (1989, 2004). If there are no conjuntive neurons or associative areas of this sort in the brain, there is no way for Barsalou's theory to cope with generality. For the sake of this paper, however, I will bracket the empirical issue altogether, and stop at the conditional claim that if conjuntive neurons exist, then a perceptual symbol systems is capable of representing general knowledge via invariant symbols.

5. Proxytypes and generality

Prinz is more committed than Barsalou to the contextual and on–line character of concepts. On his view, just like on Barsalou's view, long–term memory contains networks that organize perceptually derived knowledge about categories, which is re–activated in conceptual tasks. Reenactments are called 'proxytypes', because 'they stand as proxies for the categories they represent' in a given context (Prinz 2002: 149). Concepts are identified with proxytypes, not with networks that manage them. Thus, there are countless concepts of the same category, each one with the same causal antecedent — the category in question — but with

10. That they are part of the perceptual system is itself a debated issue, see Weiskopf (2007), and Prinz (2002: 137) for replies.

different information encoded. Prinz motivates his position on concept individuation as follows. Thoughts are occurrent states, i.e., they are processed by working memory during specific tasks (like inferring, imagining, or interpreting speech). Concepts are the components of thoughts. Therefore, concepts are occurrent representations.

This is not a convincing argument. First, the characterization of thoughts as occurrent representations is far from being a standard one — philosophers were often found to show a preference for the view that that thoughts are eternal or timeless entities, because they are truthvalue bearers. Even leaving radical Fregean views aside, drawing a line between beliefs and thoughts is disputable, and beliefs need not be occurrent (we have many non–occurrent beliefs). But if thoughts are not occurrent, concepts need not be either. Secondly, even granting Prinz that thoughts were occurrent (according to some proprietary sense of 'thought'), there is no compelling reason for concepts to be so. The so–called fallacy of distribution is at play here. The fallacy of distribution occurs when properties of a whole are attributed to its parts, for example, when it is inferred that all players of team T are good from the premise that team T is good. Team T may well feature a top–level coaching of individually low–level players. Analogously, a thought can be the temporary (selective) activation of stable representations. If it can be, then concepts need not be temporary even if thoughts are. One may also have a further, more speculative reason for being unconvinced by Prinz's argument. After all, concepts have many roles, and the claim that they are components of thoughts may not be the central one. One could live with a theory that features parts of concepts, or concept–tokens, as components of thoughts. Vice versa, it seems that a theory of concepts which didn't possess the resources to explain, say, concept combination or category induction, would be deemed inadequate to most. Which role is essential to concepts is, however, a very difficult issue to tackle, and theorists tend to have conflicting intuitions on the matter — that's why this objection against Prinz's choice is more speculative.

Why this digression about Prinz's identification of concepts with proxytypes? Because were it not for that, Prinz's theory would have been identical to Barsalou's as far as the generality issue is concerned. Prinz, like Barsalou, highlights the role of conjuntive neurons ('memory networks') in binding and grouping together proxytypes with the same informational cause. As I argued in the previous section, such structures have the resources to fulfil the abstraction and the induction function of concepts.

That said, let's assume Prinz's identification of concepts with proxytypes, and see how the proxytype theory copes with the idea that concepts encode general knowledge. As I characterized it in section 2, for conceptual knowledge to be general it suffices that it is behaviourally general, namely, that it can be applied to all instances of a category, so to perform the mental glue function required for abstraction and induction. Here, the proxytype theory takes a statistical turn. The idea is that, along with contextually generated and contextually reproduced representations, perceptual memory networks store representations of most frequently processed features of a category. These are called 'default proxytypes'. So for example my default proxytype of a dog would encode the knowledge that dogs are quadrupeds, they are furry, they are mostly friendly, they eat meat, their maximum weight is usually less than 55 kilos, you can play ball with them, they are called 'cani' in Italian and 'dogs' in English, etc. In Prinz's words, 'a default proxytype is the representation that one would token if one were asked to consider a category without being given a context'. Most frequently processed features can be highly cue–valid ones, or highly projectable ones, or perceptually salient ones etc. They differ from traditional prototypes in that they may contain linguistic information as well as theory–based knowledge, and their reference is fixed through causal links, not by confronting similarity thresholds (Prinz 2002: 154)[11].

How can proxytypes be representations of general knowledge about categories? Just like traditional prototypes do. Default proxytype support abstraction because generalizations such as would token if one were asked to consider a category without being given a context'. They also support category induction because when presented to a new member of a category, even just verbally, we automatically activate (re–enact) the default proxytype for that category. So, in Murphy's example, when I hear that my friend needs me to take care of her dog, I access my default dog–proxytype described above. Maybe not all of it, if contextual factors allow me to do some selection — for example, I may not need to recall that dogs are called '*cani*' in Italian for the task I'm required to perform.

11. For discussion see Prinz 2002, pp. 154–156.

But by and large, what makes induction possible is default prototype reenactment.

Proxytype theory, then, account for generality because they instantiate the summary representations model. We use default proxytypes most of the time[12], and they are summary representations. Keil's quotation reported in section 3 above fits in here, because most of the time, on the proxytype theory 'shared mental structures are assumed to be constant across repeated categorizations of the same set of instances and different from other categorizations'[13]. With contextual proxytypes only, the theory couldn't account for the intuitive role of concepts as representations of generality. Equipped with default representations, which encode knowledge that can be applied to all category members, Prinz's version of a situated cognition theory of concepts is no worse off than traditional competitors. But it is not utterly new either.

I'll mention a possible problem that default proxytype may have before drawing the final conclusions. Default proxytypes encode most frequently assessed knowledge, and most frequently assessed knowledge makes category variability disappear. Take dogs again, as a most (perceptually) variable category. Suppose I know that dogs come in various breeds, like German shepherds and Chihuahuas, miniature schnauzers and border collies. I've read it in books. But I have six basset–hounds at home, I live in a basset–hound populated small island, where no other dog breeds are allowed. I am also a member of the Basset–hound society that takes care of preserving the breed standards. It is likely that the contents of knowledge about dogs I employ most of the time are Basset–hounds contents of knowledge. Simply, it happens to be so. So, my default proxytype for the category of dogs strikingly resembles a Basset–hound (short–legged, sad and wide–eyed). It surely can't be the representation I employ in order to support abstraction, as in the thought that some dogs can weigh more than 45 kilos. And it can't be the representation I match with a border collie, when I finally see one in a TV commercial, in order to recognize it as a dog. This Basset–hound effect[14] casts a shadow on the adequacy of default proxytypes as representations of general knowledge, because they seem to have

12. Prinz 2002, p. 137.
13. Keil 1994, p. 169, see section 3 of this paper.
14. A similar worry has been raised about exemplar theories of concepts.

difficulties with highly variable categories. As noted in section 3 above, traditional prototype theories introduce feature weights in order to cope with that, and in general, prototypes included graded scales of features. I see no principled reason why Prinz's default proxytype couldn't include weights as well, but we are not explicitly told so. Without this indication, however, default proxytypes are subject to the basset–hound effect, which may impair their performance with highly variable categories.

6. Conclusion

In this paper I focused on the role of concepts as representations of general knowledge, which seems intuitive, but it is seldom explained. I identified two kinds of general knowledge, namely, constitutively general (possessed by all members of a category) and behaviourally general (that can be applied to all members of a category). The latter is less demanding than the former. I argued that general knowledge in either sense is necessary in order to perform abstraction and induction tasks, which are arguably core functions of concepts. Concepts can be the mental glue of cognition only if they are representations of general knowledge. Then I reviewed the ways in which theories of concepts have coped with generality. I identified two broad models. According to the first, generality is represented by invariant symbols that mark all knowledge coming from a certain category (Fodor's theory of atomic symbols is a clear example). On the second model, a summary representation encodes the properties that category instances generally have (and this is typically the case for prototype theories). Now, situated cognition theories in Barsalou's and in Prinz's version just instantiate the first and the second model respectively. Actually, Prinz's summary representations (default proxytypes) may have difficulties in explaining abstraction and induction in highly variable categories, but as I argued in the latest section, I see no good reason why, on Prinz's view, concepts couldn't be identified with memory networks rather than with proxytypes, so to instantiate the invariant symbol model. Alternatively, default proxytypes can be supplemented with weighted features. Situated cognition theories, then, face no special problem with the generality of concepts. What I have been suggesting is that it is because they are not so radically different from traditional theories as they present themselves as being.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–609.

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, 358, 1177–1187.

Barsalou, L. W., Simmons, W. K., Barbey, A., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality–specific systems. *Trends in Cognitive Sciences*, 7, 84–91.

Bloom, P. (2002). *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.

Damasio, A.R. (1989). Time–locked multiregional retroactivation: A systems–level proposal for the neural substrates of recall and recognition. *Cognition* 33, 25–62.

Damasio, A. R. (1994). *Descartes' Error*. New York: Grosset.

Fodor, J. A. (1998). *Concepts. Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

Fodor, J. A (2008) *LOT 2. The Language of Thought Revisited.* Oxford: Clarendon Press.

Keil, F. C. (1994). Explanation Based Constraints on the Acquisition of Word Meaning, *Lingua*, 92, 169–196.

Kripke, S. (1980). *Naming and Necessity*. Cambridge: Harvard University Press.

Locke, J. (1690). *An Essay Concerning Human Understanding*. P. H. Nidditch, ed. Oxford: Oxford University Press, 1979.

Markman, E. (1999). *Categorization and naming in children: problems in induction*. Cambridge: MIT Press.

Millikan, R. G. (2000). *On Clear and Confused Ideas: An Essay on Substance Concepts*. Cambridge: Cambridge University Press.

Prinz, J. J. (2002). *Furnishing the Mind*. Cambridge: MIT Press.

Rosch, E. (1973). Natural Categories. *Cognitive Psychology* 4, 328–350.

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

Smith, E. E. and Medin, D. L. (1978). *Categories and Concepts*. Cambridge: Harvard University Press.

Weiskopf, D. (2007). Concept empiricism and the vehicles of thought. *Journal of Consciousness Studies*, 14, 156–183.

Wittgenstein, L.(1953) *Philosophical Investigations*. New York, Macmillan.

Elisabetta Lalumera
University of Milan–Bicocca
Cogito Research Centre, Bologna
elisabetta.lalumera@unimib.it

Against Hybrid Theories of Concepts

EDOUARD MACHERY, SELJA SEPPÄLÄ

Abstract Psychologists of concepts' traditional assumption that there are many properties common to all concepts has been subject to devastating critiques in psychology and in the philosophy of psychology. However, it is currently unclear what approach to concepts is best suited to replace this traditional assumption. In this article, we compare two competing approaches, the Heterogeneity Hypothesis and the hybrid theories of concepts, and we present some new evidence that tentatively supports the former over the latter.

Key words Concepts, polisemy, hybrid theories of concepts, heterogeneity hypothesis, categorization, natural kind, prototypes, theories.

Psychologists working on human higher cognitive competences (e.g., categorization, induction, analogy–making, etc.) have traditionally assumed that there are numerous properties common to all concepts, and theories of concepts have attempted to characterize these common properties. There is however a growing consensus against this assumption in psychology (Medin, Lynch, & Solomon, 2000; Murphy, 2002) and in the philosophy of psychology (Machery, 2005, 2006, 2009; Piccinini & Scott, 2006; Weiskopf, 2009). Psychologists and philosophers concur that in order to explain why there are different kinds of categorization judgments, of episodes of inductive reasoning, of analogies (etc.), it is necessary to postulate that the cognitive processes that underlie our higher cognitive competences rely on *several distinct kinds* of bodies of information that are very different from one another (but see Danks, 2007). However, psychologists and philosophers disagree about how to characterize these diverse kinds of bodies of information. In this article, we compare two competing approaches: the Heterogeneity Hypothesis and the hybrid theories of concepts. We present some new evidence that tentatively supports the first approach over the second (for additional arguments, see Machery, 2009).

Here is how we will proceed. In Section 1, we review the main argument against the assumption that there are numerous properties common to all concepts. In Section 2, we describe the Heterogeneity Hypothesis and the hybrid theories of concepts. In Section 3, we present an empirical study in support of the Heterogeneity Hypothesis (Study 1). In Section 4, we present a replication of Study 1 (Study 2). In Section 5, we address four objections.

1. Against the Natural Kind Assumption

Following psychologists, we will use "concept" to refer to those bodies of information that are used by default in the processes underlying our higher cognitive competences. For instance, a concept of dog is a body of information about dogs that is used by default when we categorize something as a dog, reason inductively or deductively about dogs, draw analogies between something and dogs, understand sentences containing "dog", and so on. Psychologists interested in concepts attempt to characterize these bodies of information because their properties explain various features of our higher cognitive competences—they explain why we categorize the way we do, why we tend to draw the inductions we tend to draw, and so on.

It is common in psychology to hold (explicitly or implicitly) that concepts have many properties in common (they all store a single kind of information, they all have a single type of functional properties, they are all acquired by the same type of learning process, etc.) and that the goal of theories of concepts is to describe these properties. We call this assumption "the natural kind assumption". Recent work in psychology and in the philosophy of psychology has however challenged the natural kind assumption. If concepts have many properties in common, as the natural kind assumption would have it, then categorization judgments should have many properties in common, and similarly for our episodes of inductive and deductive reasoning, the analogies we make, etc. Thus, the natural kind assumption would be undermined if there are several types of categorization judgments that have few properties in common, several types of episodes of inductive reasoning that have few properties in common, etc. The explanatory challenge to the natural kind assumption states that there are in fact several distinct types of categorization judgments, several distinct types of episodes of inductive reasoning, and so on, and that these types are markedly different from one another.

Let us illustrate this explanatory challenge with the research on categorization. Research on categorization suggests that our decisions to categorize an object *t* as a member of the class T divide into several distinct types that have little in common. In some cases, we decide to classify *t* as a member of T if *t* possesses a sufficient number of properties among the properties that we take to be typical of the members of T. In other cases, we decide to classify *t* as a member of T if it is sufficiently similar to some salient members of T. In yet other cases, we decide to classify *t* as a member of T if it possesses the properties that we expect members of T to have in virtue of the causes that, in our light, make members of T what they are.

2. Two Answers to the Explanatory Challenge

The explanatory challenge has been answered in various ways in psychology and in philosophy. In this section, we present two answers.

2.1. The Heterogeneity Hypothesis

To meet the explanatory challenge, Machery (2005, 2009) has hypothesized that concepts divide into several types that have little in common — a view called *the Heterogeneity Hypothesis*[1]. A given class, substance, or event is typically represented by several concepts that belong to these distinct types of concepts[2]. For instance, dogs might be represented by several concepts of dog that belong to the hypothesized distinct types of concepts. In addition, the coreferential concepts (e.g., the coreferential concepts of dog) are used in distinct categorization processes, in distinct induction processes, in distinct processes of analogy–making, etc.

1. These different concept types are not meant to be merely different; they are supposed to have few (relevant) properties in common, so that there are few properties that are common to all concepts.

2. One might wonder whether these concepts are really coreferential: For instance, can a prototype and a theory have the same extension? The answer depends on how prototypes, theories, etc., refer. If, roughly speaking, they refer to their usual cause, then they are coreferential in spite of storing different kinds of information about their extension.

Because we typically possess several coreferential concepts that belong to very different types of concepts and because these concepts are used in distinct categorization processes, in distinct induction processes, in distinct processes of analogy–making (etc.), the Heterogeneity Hypothesis predicts the existence of markedly different types of categorization decisions, episodes of inductive reasoning, analogies (etc.). By contrast, recall that the natural kind assumption predicts that our categorization decisions should have many properties in common, and similarly for our inductive judgments and the analogies we make.

What are these very different types of concepts? Empirical evidence suggests that there are at least three different types of concepts: prototypes, sets of exemplars, and theories. Prototypes, exemplars, and theories store different types of information. A prototype of an *x* is a body of statistical information about the typical or cue–valid properties of *x's*[3]. An exemplar is a representation of a singular object. A concept of an *x* is a set of exemplars of *x's*. A theory of an *x* is a body of causal, nomological, or functional information about *x's*[4].

Although it might be needed to add other types of concepts to these three types in order to account for the various properties of our higher cognitive competences, Machery (2009) has argued that the necessity to do so has not been clearly shown.

2.2. Hybrid Theories of Concepts

Hybrid theories of concepts offer a different answer to the explanatory challenge. Many different types of hybrid theories of concepts have been developed since the 1970s[5], but all of them distinguish parts of concepts instead of distinguishing several types of concepts. Where the Heterogeneity Hypothesis proposes that a class (substance, event, etc.) is typically represented by several coreferential concepts (a prototype, a set of exemplars, and a theory), the hybrid theories of concepts propose

3. A property P is cue–valid for a class T if the probability that an entity belongs to T if it has P is high.

4. In the remainder of this paper, "theory" and "theoretical" will always be used in this technical sense.

5. See, e.g., Rips, Shoben, & Smith, 1973; Osherson & Smith, 1981; Keil, 1989; R. Gelman, 2004; Nosofsky et al., 1994; Keil, Carter Smith, Simons, & Levin, 1998; Anderson & Betz, 2001.

that a class (substance, event, etc.) is typically represented by a single concept that divides into several distinct parts (a hybrid concept). Hybrid theories of concepts typically hold that these parts store different types of information. For instance, one part of a concept of an *x* might store some statistical information about *x's*, while another part stores some information about specific members of the class of *x's*, and a third part some causal, nomological, or functional information about *x's* (e.g., Gelman, 2004). That is, the parts of concepts distinguished by hybrid theories of concepts might store the kind of information that is stored by the distinct concepts distinguished by the Heterogeneity Hypothesis.[6] Furthermore, hybrid theories often hold that the distinct parts that compose a given concept are used in different processes (e.g., Osherson & Smith, 1981), exactly like the coreferential concepts distinguished by the Heterogeneity Hypothesis. For instance, the parts that compose a given hybrid concept might be used in distinct categorization processes, in distinct induction processes, and so on. Because the parts of a given hybrid concept are meant to store different types of information and are often meant to be used in different processes, hybrid theories of concepts do expect categorization decisions, episodes of inductive and deductive reasoning, analogies, etc., to fall into several distinct types, in contrast to the prediction drawn from the natural kind assumption.

To illustrate the gist of the hybrid theories of concepts, let us consider the hybrid model of concepts developed by Osherson and Smith (1981). They propose that concepts are made of two parts, a core and an identification procedure. The core of a concept is supposed to be a definition (in substance, a set of properties that are deemed to be necessary and sufficient for belonging to a category), while the identification procedure is supposed to be a prototype. Thus, the core and the identification procedure of a concept are assumed to store two different types of information about its extension. Osherson and Smith also propose that some cognitive competences involve only one of these two parts. Concept composition is assumed to involve exclusively the core: When we create a complex concept about, for instance, pet fish, we use only the information stored in the core of the concepts FISH and PET. Other competences, such as categorization, involve both the definition and the prototype: Osherson and Smith propose

6. Whether it does depends on the specifics of the hybrid theory of concept.

that there are two distinct categorization processes–viz. a prototype–based process and a definition–based process.

2.3. Comparison of the Two Approaches

One might suspect that the Heterogeneity Hypothesis and the hybrid theories of concepts differ only terminologically: What the former calls "coreferential concepts" seems to correspond to what the latter view as the parts of a given concept. For instance, what the Heterogeneity Hypothesis sees as three coreferential concepts of dogs (a prototype of dogs, a set of exemplars of dogs, and a theory about dogs) seems to correspond to what hybrid theories of concepts view as the parts of a single concept of dog.

We now argue that this suspicion is wrong. Note first that hybrid theories of concepts in general are not committed to the specifics of the Heterogeneity Hypothesis. Many hybrid theories of concepts do not hypothesize that a concept of an *x* divides into three parts, which store some statistical information about *x's*, some information about specific *x's*, and some causal information about *x's* (e.g., Osherson & Smith, 1981). The hybrid theories that do not make this hypothesis clearly differ from the Heterogeneity Hypothesis and make different empirical predictions.

However, some hybrid theories of concepts do make such a hypothesis (e.g., Gelman, 2004). One might then suspect that *these* hybrid theories and the Heterogeneity Hypothesis differ only terminologically[7]. To address this issue, one needs to clarify the difference between two bodies of information being *two distinct concepts* and two bodies of information being *two parts of a given concept*. That is, what does it mean to say that, e.g., a body of information about the typical properties of *x's* and a body of information about the causal properties of *x's* are two parts of a single concept rather than two distinct concepts?

Proponents of hybrid theories of concepts have not given an explicit answer to this question. It is however possible to abstract the distinction

7. In the remainder of this paper, the expression "hybrid theories of concepts" refers exclusively to those hybrid theories of concepts that would postulate that concepts have three parts: one that stores some prototypical information, one that stores some information about specific individuals, and one that stores some theoretical information. The other hybrid theories are undermined by the fact that they cannot explain the existing diversity of our categorization judgments, of our episodes of inductive reasoning, etc.

of interest from the models of hybrid concepts they have developed. If two coreferential bodies of information fulfill the two following necessary conditions, they are parts of a given concept:

1. These bodies of information are necessarily *linked* to each other: Using one of them to, e.g., categorize, enables me to use the other one for other purposes (e.g., to reason deductively or inductively).
2. These bodies of information are *coordinated*: They do not produce conflicting outcomes that are irreducible (for instance, inconsistent categorization judgments that are taken to be equally authoritative).

If one of these two conditions is not fulfilled, two coreferential bodies of information are two distinct concepts.

Let us illustrate these two conditions for conceptual parthood by means of simple examples. Take my belief that water is essentially made of H_2O and my belief that water is typically a transparent, tasteless, odorless liquid that is found in bottles or in lakes. According to the Heterogeneity Hypothesis, the first belief is the type of beliefs theories are supposed to be made of, and the second belief is the kind of beliefs prototypes are supposed to be made of. If these two beliefs are parts of a single concept of water, then classifying a substance as being water because it is (say) a transparent liquid in a bottle enables me to use the first belief to conclude that this liquid is made of H_2O (condition 1).

Furthermore, suppose these two beliefs do not produce irreducible conflicting judgments (condition 2). For instance, it does not happen that I judge *both* that a given liquid *is* water because it is transparent, tasteless, and odorless *and* that it *is not* water because it is not made of H_2O. How should the belief that water is essentially made of H_2O and the belief that water is typically a transparent, tasteless, odorless liquid that is found in bottles or in lakes be organized so as to ensure that condition 2 is met? Irreducible conflicting judgments might be avoided in at least two different ways. First, these two beliefs might be used to produce different types of judgments. If one belief is used to classify, but not to draw inductive inferences, while the other is used to draw inductive inferences, but not to classify, they would not produce conflicting judgments. Second, if both beliefs can be used to produce the same type of judgment (e.g., categorization judgments), the judgments reached on the basis of one of these two beliefs, but not the judgments reached on the basis the other belief, might be defeasible. The judgments made on the basis of the belief that water is a transparent, tasteless, odorless liquid might always be defeated if they contradict the judgments reached on the basis of the belief that water is necessarily made of H_2O. Because one of the judgments is defeated when there is a conflict, our two beliefs about water would not result in irreducible conflicting judgments. They would be coordinated.

Let us consider now a real hybrid model of concepts to illustrate the second necessary condition of conceptual parthood (coordination). As we saw above, Osherson and Smith (1981) propose that we can categorize by means of two distinct cognitive processes. One of them relies on prototypes (Osherson and Smith's identification procedure), while the other one relies on definitions (Osherson and Smith's core). In their model, the prototype of an *x* and the definition of an *x* that are meant to constitute the hybrid concept of an *x* are coordinated. Prototype–based categorization judgments about *x* are defeated if they contradict the judgments reached on the basis of the hypothesized definitional part of the concept of *x*. If people have such a hybrid concept of *x*, then, they should not hold irreducible conflicting categorization judgments.

In contrast to the hybrid theories of concepts, the Heterogeneity Hypothesis assumes that the coreferential bodies of information do not meet either condition 1 or condition 2. They are distinct concepts, rather than distinct parts of a single concept. To use our toy example, the belief that water is necessarily H_2O and the belief that water is typically a tasteless and transparent liquid might not be linked: When I categorize some substance as being water because it is a tasteless and transparent liquid, I might not be *ipso facto* able to conclude that it is made of molecules of H_2O. Alternatively, these two beliefs might not be coordinated. They might not be organized in such a way that when they produce conflicting judgments, one of these two judgments is defeated. As a result, people might hold irreducible conflicting judgments about water. In what follows, we will assume that what really distinguishes distinct coreferential concepts from distinct parts of a given concept is that the latter, but not the former, are coordinated[8]. If two bodies of information are not coordinated, we will say that they are *independent*.

---

8. The main reason is that our coreferential bodies of information are typically linked: Deciding that something is, e.g., water because it is a transparent, odorless liquid typically enables me to judge that it is probably made of H_2O.

To summarize, while the Heterogeneity Hypothesis and the hybrid theories of concepts can both explain why there are several distinct types of categorization judgments, episodes of inductive reasoning, etc., they are not mere terminological variants. Thus, one wants to know which approach best characterizes the diversity of the kinds of bodies of information used to categorize, draw analogies, etc. Below, we present an exploratory experimental study that tentatively supports the Heterogeneity Hypothesis.

3. Study 1: Conflicting Categorization Judgment

3.1. The Polysemy Hypothesis

Suppose that John Doe has several concepts of dog, as the Heterogeneity Hypothesis would have it. Then, "dog" in John Doe's idiolect might well be polysemous: "dog" might have different meanings because it might express John Doe's distinct concepts of dog. We call the hypothesis that words are so polysemous *the polysemy hypothesis*[9].

What kind of evidence could support the polysemy hypothesis? Suppose that "dog" expresses two concepts of dog, DOG_1 and DOG_2. Suppose further, as the Heterogeneity Hypothesis would have it, that one of these two concepts (DOG_1) is a prototype, while the other one (DOG_2) is a theory. According to prototype theories of concepts and categorization, we decide that an object is an *x* if and only if it is sufficiently similar to the prototype of the class of *x's* (e.g., Hampton, 1998). By contrast, according to most theory theories of concepts and categorization, our categorization decisions do not depend on whether the object is similar to the relevant prototype (e.g., Rips, 1989). There might then be some objects that are sufficiently similar to the prototype of dogs to be judged to be dogs if we categorize by means of the prototype of dogs, but that are not judged to be dogs if we categorize by means of the theory of dogs. Because DOG_1 and DOG_2 are two distinct concepts (and not two parts of a concept), they

9. The Heterogeneity Hypothesis does not *entail* the polysemy hypothesis. Even if we have several concepts of dog, it could be that only one of them is linguistically expressed. But the polysemy hypothesis is consistent with, and fits the gist of, the Heterogeneity Hypothesis.

are not coordinated[10]. People should thus be disposed to hold conflicting judgments about whether these objects are dogs (see Section 2.3). Then, if the polysemy hypothesis is correct, they should be willing to assent to apparently contradictory sentences about whether some objects are dogs[11]. That is, for some objects, they should be willing to assent to the sentences "This is a dog" and "This is not a dog".

The hybrid theories of concepts do not make this prediction. The parts of a hybrid concept are supposed to be coordinated: They do not produce irreducible conflicting judgments. If we decide that an object is an *x* when we categorize by means of a part of a concept (say, the part of a concept that stores some information about the typical properties of objects) and that it is not an *x* when we categorize by means of another part (say, the part of a concept that stores some causal, nomological, or functional information), one of these two judgments (typically the prototype–based judgment) is defeated. As a result, people should not hold irreducible conflicting judgments about whether some object is an *x*, and they should not be disposed to assent to apparently contradictory sentences about whether some object is an *x*[12].

To summarize, the polysemy hypothesis derived from the Heterogeneity Hypothesis predicts that for some objects that are sufficiently similar to the prototype of a class, but that are not members of this class according to our theory of this class, we should be disposed to assent to apparently contradictory sentences about their membership in this class. Similarly, the polysemy hypothesis predicts that for some objects that are not similar at all to the prototype of a class, but that are members of this class according to our theory of this class, we should be disposed to assent to apparently contradictory sentences about their membership in this class. Hybrid theories predict that in both cases people should be reluctant to assent to seemingly contradictory sentences.

---

10. Remember that we are assuming that what really distinguishes distinct coreferential concepts from distinct parts of a given concept is that the latter, but not the former, are coordinated.

11. If the polysemy hypothesis is right, the contradictions are obviously *only apparent* because people associate a given word with two different concepts. For instance, if people are willing to assent to "Whales are not fish" *and* to "Whales are fish", they are not really contradicting themselves because "fish" expresses two different concepts in these two sentences.

12. The natural kind assumption makes the same prediction.

Before examining how these predictions fare empirically, we want to highlight two caveats. First, even if the polysemy hypothesis is true, we do not expect everybody to be disposed to assent to apparently contradictory sentences, for some people might be reluctant to endorse apparent contradictions. The polysemy hypothesis would be supported if a *substantial* proportion of individuals are willing to assent to the kind of apparently contradictory sentences described above. Second, even if people typically have several coreferential concepts for classes, substances, etc., it is possible that, for some people, some predicates express only one of their coreferential concepts and that the other coreferential concepts are not lexicalized. The Heterogeneity Hypothesis, *but not the polysemy hypothesis*, would then be true of these people. Thus, finding that, contrary to what the polysemy hypothesis predicts, some people do not assent to apparent contradictions, such as "Whales are fish" and "Whales are not fish", is consistent with the Heterogeneity Hypothesis. By contrast, the hybrid theories of concepts discussed in this article unequivocally predict that people should not assent to this type of contradictions.

3.2. Participants and Materials

Participants were 24 individuals taking classes at the University of Pittsburgh. Two participants were excluded because they did not fill the survey properly, resulting in a sample of 22 participants (mean age = 23; range: 19–45; 45% males). Almost all participants were majoring in a scientific discipline or in history and philosophy of science.

In classroom settings, participants were presented with a survey composed of 9 pairs of sentences. Each pair was written on a separate page. All the pages were similarly organized. Participants read the following instructions on the top of each page:

> On a scale of 1 to 7, "1" indicating that you totally disagree and "7" indicating that you totally agree, to which extent do you agree with the following two claims? Remember that you might agree with both claims, with only one of the two claims, or with none of them.

A first sentence was written below these instructions. For instance, the first sentence of the first pair was "In a sense, tomatoes are vegetables". A 7–point scale, anchored at 1 with "clearly disagree", at 4 with "not sure", and at 7 with "clearly agree", followed this sentence. The negation of the first sentence was written below this scale[13]. For instance, the second sentence of the first pair was "In a sense, tomatoes are not vegetables". The same 7–point scale followed this second sentence. Participants could leave a comment after each answer.

All the sentences expressed (positive or negative) classification judgments: They asked whether a class was included in another class. The 9 pairs of sentences consisted of 6 target pairs and 3 control pairs. The 6 target pairs were constructed according to one of the two following principles: (1) in 4 pairs (A, D, E, and H), we assumed that the members of the extension of the first predicate of the sentence (e.g., "tomatoes") were similar to the hypothesized prototype expressed by the second predicate (e.g., "vegetables"), but did not belong to the extension of this predicate according to the theory we hypothesized it also expresses; (2) in the 2 remaining pairs (B and G), we assumed that the members of the extension of the first predicate (e.g., "penguins") were dissimilar to the hypothesized prototype expressed by the second predicate (e.g., "birds"), but did belong to the extension of this predicate according to the theory we hypothesized it also expresses[14]. Let us illustrate construction principle 1 with pair A. We assumed that tomatoes are similar to the prototype of vegetables because they have many of the typical properties of vegetables. At the same time, many people believe that tomatoes are essentially fruits, not vegetables. If the polysemy hypothesis is correct, then participants (or a substantial proportion of them) should be willing to assent both to "In a sense, tomatoes are vegetables" and "In a sense, tomatoes are not vegetables". Let us illustrate now construction principle 2 with pair B. We assumed that penguins are dissimilar to the prototype of birds because they have few of the typical properties of birds (they do

13. Thus, the first sentence of each pair was affirmative, while the second was negative.

14. These two construction principles assume the correction of our intuitions about what theoretical concept and what prototype are expressed by the predicates we used. We are quite confident about the prototypes we hypothesized, but we recognize that some hypothesized theoretical concepts might be controversial. For instance, we assumed that it follows from the hypothesized theory of pieces of furniture that musical instruments are not pieces of furniture (because according this hypothesized theory, pieces of furniture and musical instruments are exclusive categories of artifacts). We also assumed that the folk theory of sports is such that to be a sport, it is sufficient to have competitions and competitors. As we will see below, some participants did not share our intuitions. Their answer was removed from the data set (see Section 3.3 for explanation).

not fly) and because they have several properties that are atypical of birds (they swim). At the same time, many people believe that penguins are essentially birds. If the polysemy hypothesis is correct, then participants (or a substantial proportion among them) should be willing to assent both to "In a sense, penguins are birds" and "In a sense, penguins are not birds". We call "theoretical sentences" the sentences that we hypothesized would be judged true on the basis of the theories of the classes at hand (e.g., "In a sense, tomatoes are not vegetables" and "In a sense, penguins are birds") and "prototypical sentences" those sentences that we hypothesized would be judged true on the basis of the prototypes of the classes at hand (e.g., "In a sense, tomatoes are vegetables" and "In a sense, penguins are not birds").

Every second target pair was followed by a control pair (C, F, and I). We assumed that participants would judge the first sentence of a control pair to be clearly true and the second sentence to be clearly false, whatever concept (prototype, theory, etc.) they associate with the second predicate of the sentence. For instance, pair C was made of the two following sentences: (C1) "In a sense, lions are animals" and (C2) "In a sense, lions are not animals". Because lions are animals according to our theoretical beliefs about animals and are typical animals, people should judge C1 true and C2 false whatever concept they associate with "lion".

The 9 pairs are presented in Table 1.

Pair	First sentence on a given page	Second sentence on a given page
A	In a sense, tomatoes are vegetables	In a sense, tomatoes are not vegetables
B	In a sense, penguins are birds	In a sense, penguins are not birds
C	In a sense, lions are animals	In a sense, lions are not animals
D	In a sense, whales are fish	In a sense, whales are not fish
E	In a sense, a piano is a piece of furniture	In a sense, a piano is not a piece of furniture
F	In a sense, a triangle is a geometric figure	In a sense, a triangle is not a geometric figure
G	In a sense, chess is a sport	In a sense, chess is not a sport
H	In a sense, zombies are alive	In a sense, zombies are not alive
I	In a sense, a hammer is a tool	In a sense, a hammer is not a tool

Table 1. Target and Control Sentences (Control Pairs in Grey Shading, Theoretical Sentences in Italics, Prototypical Sentences in Regular Fonts).

In our target sentences, we used words expressing concepts that belong to different domains (plants, animals, artifacts, human activities).

This is meant to reflect the domain–generality of the Heterogeneity Hypothesis: It is supposed to apply to all conceptual domains.

Finally, the 9 pairs of sentences were followed by a short biographic questionnaire.

3.3. Results

To analyze our results, we computed the percentage of participants who gave an answer superior or equal to 4 for both sentences.[15] For each target pair, we eliminated those participants who gave an incorrect answer to the theoretical sentence.[16] For instance, in pair A, we did not take into account the answers of those participants who answered negatively (an answer lower than 4) to the sentence "In a sense, tomatoes are not vegetables" (A2), when we computed the scores for our two dependent measures. The reason is that we are interested in the following question: When people know that a tomato does not fulfill the theory of a vegetable, are they still willing to assent to the sentence, "In a sense, tomatoes are vegetables"? To examine this question, one should only examine the answers of those participants who know that a tomato is (in a sense!) not a vegetable. Generally, in pairs A, D, E, and H, we were only interested in the answers of those participants who have acquired the belief that although *x's* look like *y's*, they are *z's*; we wanted to know whether they'd also be willing to agree that, in a sense, *x's* are *y's*. Similarly, in pair B, we did not take into account the answers of those participants who answered negatively (an answer lower than 4) to the sentence "In a sense, penguins are birds" (B1). The reason is that we are interested in the following question: When people know that a penguin does fulfill the theory of a bird, are they still willing to assent to the sentence, "In a sense, penguins are not birds"? To examine this question, one should only examine the answers of those participants who know that a penguin is (in a sense!) a

15. One might object that it is inappropriate to take a response of 4 as agreeing with a sentence, and one might contend that only answers superior to 4 should be examined. However, we defend our scoring procedure on the following grounds: If one holds that whales are not a fish, failing to reject the claim that whales are a fish (which is what one would do by answering 4 or higher) is unexpected if one believes that theory–based judgments should defeat prototype–based judgments when these disagree.

16. In the tables below, we indicate the percentage of data points that were eliminated according to this coding principle.

bird. Generally, in pairs B and G, we were only interested in the answers of those participants who have acquired the belief that although *x's* do not look like *y's*, they are *y's*; we wanted to know whether they'd be also willing to agree that, in a sense, *x's* are not *y's*. The results for each pair are presented in Table 2, and the averaged results across the target and control pairs in Table 3.

	Percentage of agreement with both sentences (≥ 4)	Proportion of eliminated data points
A	28.6%	36.3%
B	9.5%	4.5%
C	0%	
D	21.1%	13.6%
E	23.1%	40.9%
F	4.17%	
G	25%	45.5%
H	60%	9.1%
I	4.17%	

Table 2. Results of Study 1 (Control Pairs in Grey Shading).

	Average percentage of agreement with both sentences (≥ 4)
Target pairs	27.9%
Control pairs	2.78%

Table 3. Averaged Results of Study 1.

3.4. Analysis and Discussion

A larger proportion of participants were willing to assent to a seeming contradiction for our target pairs (mean = 27.9%; range: 9.5%–60%) than for our control pairs (mean = 2.78%; range: 0%–4.17%)[17]. Thus, as we had predicted, some people are willing to assent to two sentences that appear to be contradictory when one sentence is judged to be true on the

17. Given the complexity of our coding scheme, we did not analyze our data statistically. We note that the qualitative analysis of the data presented below reveals clear trends. We are also relatively confident that the patterns we found are not merely due to chance since we replicated our study with two additional samples (Section 4).

basis of a prototype of a class while the other is judged to be true on the basis of the theory of the same class. This finding provides some tentative evidence for the Heterogeneity Hypothesis over the hybrid theories of concepts, for, according to the latter, the parts of concepts should *not* yield irreducible conflicting judgments, while the Heterogeneity Hypothesis suggests that people might often associate several concepts with a single lexeme and might thus sometimes be willing to assent to apparent contradictions. While some people seem reluctant to assent to apparent contradictions, we simply predicted, as we explained above, that a substantial proportion of subjects would be willing to assent to apparently contradictory sentences.

We also found that the pattern of answers just described holds for every pair. For most target sentences (A, D, E, and G), around 25% of participants were willing to assent to a seeming contradiction—a substantial proportion. Two pairs stood out: Only 10% of participants were willing to assert that in a sense penguins are not birds in addition to asserting that in a sense they are birds. By contrast, 60% of participants were willing to assert both that in a sense zombies are alive and that in a sense they are not alive. The significance of this variation is unclear.

It is also noteworthy that we did not find any clear variation across conceptual domains. Furthermore, we did not find any clear difference between the sentences constructed according to construction principle 1 (A, D, E, H) and the sentences constructed according to construction principle 2 (B and G). However, the small number of target sentences prevents us from drawing any confident conclusion from these two results.

We are aware of the limits of this study, and we acknowledge that it provides only tentative support for the Heterogeneity Hypothesis over the hybrid theories of concepts. We used a small number of target sentences, which casts doubts on whether our results can be generalized. Most important, the proportion of participants willing to endorse both sentences within each pair was lower than expected. We hypothesized that this might be due to the scientific background of our participants. Scientists might be trained to associate lexemes with a single concept (particularly with a theoretical one) even if they have several concepts. To test this hypothesis, we replicated our first study with a different sample.

4. Study 2: Replication

4.1. Participants and Materials

Study 2 was identical to study 1 except that it was conducted on–line in order to reach a more diverse group of participants. 52 participants correctly completed the survey, which was in English. We divided the participants into 2 groups—native speakers of English (N = 22, mean age = 31; age range: 20–59; 37% males) and non–native speakers of English (N = 30; mean age = 36; age range: 22–66; 44% males). Around 40% of the sample of native speakers of English and around 60% of the sample of non–native speakers reported having a non–scientific activity.

4.2. Results

We used the same method of data analysis as in Study 1 (Section 3.3). We report the results of our two samples separately. Focusing first on the non–native speakers, the results for each sentence pair are presented in Table 4, and the averaged results across the target and control pairs in Table 5.

	Percentage of agreement with both sentences (≥ 4)	Proportion of eliminated data points
A	45%	26.7%
B	20%	16.7%
C	0%	
D	28%	16.7%
E	30%	33.3%
F	38%	
G	58,8%	43.3%
H	64%	16.7%
I	10%	

Table 4. Results of Study 2, Sample of Non–Native Speakers (Control Pairs in Grey Shading).

	Average percentage of agreement with both sentences (≥ 4)
Target pairs	41.0%
Control pairs	4.6%

Table 5. Averaged Results of Study 2, Sample of Non–Native Speakers.

Turning now to the native speakers, the results for each pair are presented in Table 6, and the averaged results across the target and control pairs in Table 7.

	Percentage of agreement with both sentences (≥ 4)	Proportion of eliminated data points
A	80%	9.1%
B	36%	0%
C	4.5%	
D	31.8%	0%
E	90%	9.10%
F	25%	
G	72.7%	50%
H	73.7%	13.60%
I	19%	

Table 6. Results of Study 2, Sample of Native Speakers (Control Pairs in Grey Shading).

	Average percentage of agreement with both sentences (≥ 4)
Target pairs	64.0%
Control pairs	16.2%

Table 7. Averaged Results of Study 2, Sample of Native Speakers.

4.3. Analysis and Discussion

The results of Study 2 are by and large consistent with our previous results and provide further evidence in support of the polysemy hypothesis derived from the Heterogeneity Hypothesis. This is particularly the case when one focuses on the averaged results. A larger proportion of non–native speakers were willing to assent to a seeming contradiction for our target pairs (mean = 41.0%; range: 20%–64%) than for our control pairs (mean = 4.6%; range: 0%–10%). Similarly, a larger proportion of native speakers were willing to assent to a seeming contradiction for our target pairs (mean = 64%; range: 31.8%–90%) than for our control pairs (mean = 16.2%; range: 4.5%–25%). The same pattern of answers also holds when one considers the answers to each pair in the two samples of Study 2.

Furthermore, as we had expected, when the background of our participants was less scientific, the percentage of answers consistent with the polysemy hypothesis substantially increased (from 28% to 41% and 64%). However, since no statistical analysis was conducted, this result should be treated with caution.

Let us now compare the three samples (Figure 1). Overall, the answers given by the non–native speakers in Study 2 were similar to the answers given by the participants in Study 1 (who were all native speakers). The answers given by the native speakers in Study 2 differed somewhat from the answers given by the two other samples. Native speakers in Study 2 were much more likely than the non–native speakers of Study 2 or the native speakers of Study 1 to assent to both sentences of pairs A and E. Otherwise, their answers were similar to the answers given by the other participants.

In all three samples, people were reluctant to agree with both sentences of pair B, exactly as we had found in Study 1. Most participants also tended to agree with both sentences of pair H, as we had found in Study 1. We cannot draw any confident conclusion about the existence of differences across conceptual domains, even though the proportion of agreements with both sentences seems somewhat lower for the two pairs involving animals (B and D) if one averages across the three samples.

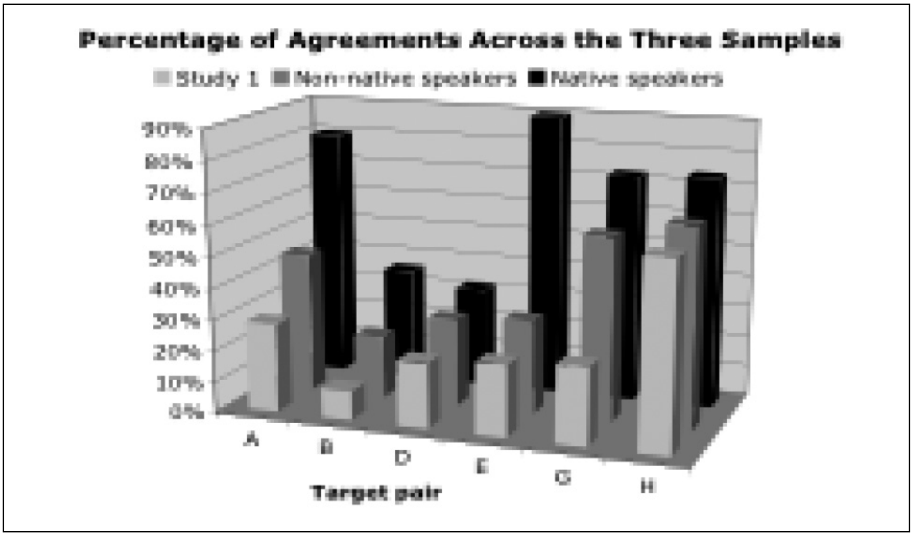

Figure 1. Percentage of Agreements to Both Sentences for each Target Pair across the Three Samples.

To conclude, our findings tentatively support the Heterogeneity Hypothesis over the hybrid theories of concepts. For a substantial proportion of people, the coreferential bodies of information do not seem to be coordinated: They are not organized so as to prevent irreducible conflicting judgments. On the contrary, coreferential bodies of information, e.g., our bodies of information about vegetables, can lead to conflicting judgments of categorization.

That said, we acknowledge again the tentative nature of the evidence presented in this article. We view this evidence as raising a challenge for the hybrid theories of concepts rather than as a fatal objection.

5. Objections and Replies

In the last section of this article, we consider the main objections to the empirical argument developed in this article.

5.1. "In a Sense"

One could first object that using the hedge "in a sense" at the beginning of each sentence might have led participants to reinterpret the sentences so as to agree with both sentences. Alternatively, "in a sense" might invite participants to answer about the pragmatic felicity of the sentences we used rather than their truth. Thus, showing that people agree to apparently contradictory sentences that do not include this hedge, e.g., "whales are fish" and "whales are not fish", would be stronger evidence against hybrid theories.

We agree that replicating our study without "in a sense" would strengthen our results, and we plan to examine this question in further work. However, we doubt that participants reinterpreted the sentences or merely agreed to their pragmatic felicity. The control sentences suggest that merely adding "in a sense" is not sufficient for people to agree with apparently contradictory sentences, while the explanations participants volunteered (some of which are given in Section 5.2) suggest that their answers reflect their opinions as to the truth–values of the relevant sentences, literally understood.

5.2. Mere Metaphors?

One could also object that one of the two sentences in a given pair is interpreted literally, while the other one is interpreted metaphorically:

When people assert that tomatoes are not vegetables, they mean it literally; by contrast, when they assert that tomatoes are vegetables, they really mean that tomatoes are *like* vegetables in various respects. According to this argument, participants did not hold conflicting judgments because they asserted on the one hand that tomatoes were not vegetables and, on the other hand, that they were *like* vegetables. If they don't hold conflicting judgments, the objection concludes, our findings provide no evidence against the hybrid theories of concepts.

We have two mutually consistent replies to this objection. First, when one looks at the comments left by participants, one finds little evidence in support of the idea that one of the two conflicting sentences is interpreted metaphorically. Here are some of the comments left by those participants in Study 2 who gave an answer higher than or equal to 4 for the sentence "In a sense, tomatoes are vegetables":

1. They're grouped with vegetables on menus.
2. They have seeds, so they are a fruit — but the human use we make of them in cooking is more similar to the ways we cook other vegetables.
3. I believe they are fruit, but we eat them like vegetables, so in a sense, tomatoes are vegetables.
4. Botanically they are fruits, but they are not sweet, and are used primarily in salty dishes and in conjunction with other vegetables.
5. We often associate them to vegetables, by the way we eat them (salads, main dish, instead of desserts).

These comments fit strikingly well with the polysemy hypothesis derived from the Heterogeneity Hypothesis. Participants seem to assert that tomatoes are not vegetables because they hold that tomatoes are fruits and that fruits and vegetables are exclusive biological categories. They also seem to assert that tomatoes are nonetheless vegetables because they have many typical properties of vegetables. That is, one judgment seems driven by a theoretical concept of vegetables (one that involves a belief in a biological classification made of exclusive categories and discovered by science), while the other seems driven by a prototype of vegetables. In addition, it is noteworthy that in Study 2 (for which we have most comments), we found no reference whatsoever to a metaphorical reading of the target sentences.

We concede that our first reply is not decisive. To examine the matter more directly, we conducted a third experimental study. Study 3 was conducted on–line in English. 109 participants took part in the experiment. 5 participants did not complete the survey properly and were removed from the sample, resulting in a sample of 63 native speakers (mean age: 30; range: 18–52; 59% males) and a sample of 41 non–native speakers (mean age: 26; range: 18–37; 59% males)[18].

Participants were presented with the following text:

Figurative language uses "figures of speech", like metaphors, which differ from ***literal*** speech. For instance, when a sports announcer says "Shaq is ***on fire*** tonight", this is an instance of figurative language. If Shaq were really on fire, he would suffer burns. What the announcer is saying with his figurative sentence, "Shaq is on fire", is that Shaq is performing extremely well.

They were then given the following instructions:

For each of the following sentences, we have underlined a word, and we want you to indicate the extent to which you think this word is used literally or figuratively on a scale of 1 to 7, "1" indicating that you think it is used figuratively and "7" indicating that you think it is used literally.

These instructions were followed by 9 sentences (Table 8), each of which was followed by a 7–point scale, anchored at 1 with "literal" and at 7 with "figurative". 4 of these 9 sentences were control sentences (in grey in Table 8), meant to check whether participants had understood what a metaphorical sentence was. In addition to these control sentences, we included 5 of the 12 target sentences used in Surveys 1 and 2. For each sentence, participants had the possibility to explain their answer.

	Sentences	Native speakers	Non–native speakers
S1	In a sense, lions are **kings** among animals	6.19 (1.34)	5.88 (1.87)
S2	In a sense, tomatoes are **vegetables**	3.29 (2.09)	2.78 (2.30)
S3	In a sense, elephants are **animals**	1.13 (0.77)	1.37 (1.37)
S4	In a sense, a plane is a **bird**	6.27 (1.26)	6.34 (1.48)

18. We are grateful to Shaun Nichols for having shared with us the prompts used in this study.

	Sentences	Native speakers	Non–native speakers
S5	In a sense, chess is a **sport**	3.05 (1.77)	3.51 (2.42)
S6	In a sense, whales are **fish**	4.87 (1.84)	4.51 (2.19)
S7	In a sense, some people are **pigs**	6.06 (1.06)	6.12 (1.82)
S8	In a sense, zombies are **alive**	4.21 (1.85)	4.68 (2.11)
S9	In a sense, a piano is **a piece of furniture**	2.87 (2.00)	3.32 (2.34)

Table 8. Means and Standard Deviations (in Parentheses) for Study 3 (Control Sentences in Grey Shading).

	Native speakers	Non–native speakers
Control sentences (literal)	1.13 (0.77)	1.37 (1.37)
Control sentences (metaphorical)	6.17 (1.22)	6.11 (1.72)
Target sentences	3.66 (1.91)	3.76 (2.27)

Table 9. Averaged Means and Standard Deviations (in Parentheses) of Study 3.

	S1	S2	S3	S4	S5	S6	S7	S8	S9
S1		**	**	ns	**	**	ns	**	**
S2			**	**	ns	**	**	ns	ns
S3				**	**	**	**	**	**
S4					**	**	ns	**	**
S5						**	**	*	ns
S6							**	ns	**
S7								**	**
S8									**

ns = non–significant; * < .05; ** < .005

Table 10. Pairwise Comparison for the Sample of Native Speakers (Control Metaphorical Sentences in Lighter Grey, Control Literal Sentence in Darker Grey).

The mean answers and standard deviations for each sentence are presented in Table 8, and the average answers for the control sentences and the target sentences are presented in Table 9. A pairwise comparison of the 9 sentences for the sample of native speakers (adjusted for multiple comparisons with the Bonferroni procedure) is reported in Table 10[19].

19. The results for the sample of non–native speakers were by and large similar except for the sentence "In a sense, lions are kings among animals", which received a lower rating.

It is apparent from Tables 8 and 9 and confirmed by Table 10 that the 5 target sentences used in Studies 1 and 2 are viewed as much more literal than the control metaphorical sentences. It is thus unlikely that people answered the way they did in Studies 1 and 2 because they treated some sentences—viz. "In a sense, tomatoes are vegetables" or "In a sense, pianos are pieces of furniture"—as metaphorical.

On the other hand, it is true that the target sentences in Study 3 are also viewed as much less literal than the control literal sentence (S3) and that some target sentences—particularly, "In a sense, whales are fish"—received a rather high mean rating. One might wonder whether this does not provide some evidence that participants who agreed with both sentences for the target pairs in Studies 1 and 2 read sentences such as "In a sense, whales are fish" as being metaphorical.

In reply, we note that, when we looked at the comments left in Study 3, we found that some participants thought that the sentences they took to be false were metaphorical. They might have reasoned that if someone were to utter these (in their mind, false) sentences, she would have to utter them metaphorically. Thus, several participants who noted in their comment that the sentence "In a sense, whales are fish" was false gave a rating of 7—viz. metaphorical—on our scale:

6. The common sense notion of whale includes the concept 'mammal'.
7. Whales are mammals, where fish are not — they may both be in the ocean, but they are different organisms, in the same way that monkeys are different from termites, though both hang around in jungles.
8. Whales are mammals.

Participants made the same kind of comments for other target sentences too. For instance, a participant who answered "7" for the sentence "In a sense, chess is a sport" wrote that "Chess is a sport for some players, but it is an Arabian game", while a participant who answered "7" for the sentence "In a sense, tomatoes are vegetables" wrote that "tomatoes are not vegetables, technically I think they're fruit", and a participant who answered "7" for the sentence "In a sense, zombies are alive" wrote that "they are literally dead, but don't act like dead people, they move about for a start, which would lead people to say they are alive". Such comments suggest that when people are disposed to view the assertion of a sentence such as "In a sense, whales are fish" as metaphorical, they clearly express their opinion that it is false. We thus doubt that those participants in Studies 1 and 2 who answered that "In a sense, whales are fish" and other relevant sentences were true read them metaphorically.

5.3. A Single–Concept Explanation?

One could argue that in order to account for our findings, there is no need to postulate that predicates, such as "vegetable" or "fish", express several coreferential concepts. To see why, consider the sentence G1 "In a sense, chess is a sport". Chess has several typical properties of sports (there are chess competitions, it involves training, etc.), but it also lacks several other typical properties of sports (it does not involve physical effort, it is not shown on TV, etc.). Suppose that people have a single concept of sport and that this concept is a prototype[20]. Then, people could find sentences G1 and G2 to be both true. They would find G1 true because the similarity between chess and people's hypothesized prototype of sport barely reaches the threshold needed to judge that something is a sport. At the same time, they would find G2 true because the similarity between chess and people's hypothesized prototype of sport is not high enough to assert confidently that chess is a sport. It is thus possible to explain why people might assent to both "In a sense, chess is a sport" and "In a sense, chess is not a sport" by assuming that a word like "chess" expresses a single concept. If this objection were correct, then our findings might be problematic for the hybrid theories of concepts, but would fail to support the Heterogeneity Hypothesis.

We have three mutually consistent replies to this objection. First, it does not seem to account very well for the answers to some pairs of sentences (particularly, A, B, and D). Consider for instance A2, "In a sense, tomatoes are not vegetables". The objection proposes that those participants who agree with A2 do so because they judge that tomatoes lack some typical properties of vegetables. However, we doubt that this is the correct explanation because tomatoes have all the typical properties of

20. We could run the same argument by supposing that people have a single concept of sport and that this concept is a set of exemplars of sports. But the argument would not work if we hypothesized that people have a single concept of sports and that this concept is a theory of sports. For, according to theory–based models of categorization, people's decision that something falls under a concept C does not depend on this thing having many properties that are typical of the members of the extension of C.

vegetables: They are eaten like vegetables, they taste like vegetables, and they are sold with vegetables.

Second, the comments left by participants do not support the hypothesis that they answered to all target sentences on the basis of prototypes. Consider two comments left for A2:

9. According to the scientific definition, they are a fruit.
10. I believe tomatoes are fruits, but we eat them like vegetables. But since I think they are a fruit, then, in a sense, tomatoes are not vegetables.

Participants seemed often to deny that tomatoes are vegetables because they are scientifically classified as fruits (or because they have allegedly essential properties of fruits) and because fruits and vegetables are scientifically exclusive categories.

In reply, our critic might concede that her objection applies best to pairs E, G, and H, for just like chess has some typical properties of sports, but not others, pianos have some typical properties of pieces of furniture, but lack others, and zombies have some typical properties of living creatures, but lack others. Our critic might then insist that her objection still undermines our claim that our findings support the Heterogeneity Hypothesis because half the sentences we used do not provide evidence for a multiplicity of coreferential concepts. To see why the objection under consideration, even hedged in this way, ought to be resisted, we turn to our last reply.

This objection supposes that the objects that are classified in the target sentences are borderline cases—that is, objects that neither clearly belong to nor clearly do not belong to the extension of the relevant predicates: Because they (e.g., tomatoes) have some typical properties, but not all, of other objects (e.g., vegetables), their membership in the relevant extension (e.g., in the extension of "vegetables") is unclear. Suppose now that the answer given on the scale reflects whether the objects to be classified are clear instances of the category at hand. That is, if participants think that the objects to be classified are clear instances, they answer "7"; if they think the objects are unclear instances, they answer "4". If this assumption is granted, then our critic is committed to asserting that participants should answer "4" or "5" for both sentences in each pair. To test this hypothesis, we determined the proportion of participants in Studies 1 and 2 who answered 4–4, 4–5, 5–4, and 5–5 among those who agreed with both sentences in each pair. Table 11 reports our results.

Pair	Study 1	Study 2 (Native Speakers)	Study 2 (Non–Native Speakers)
A	0%	12.5%	10%
B	0%	37.5%	20%
D	75%	28.6%	25%
E	0%	11.1%	0%
G	20%	0%	37.5%
H	66.6%	7.1%	50%

Table 11. Proportion of Participants who Answered 4, 5, or 4 and 5 to the Sentences in each Pair.

Our results do not support the explanation suggested by our critic. For all proportions reported in Table 11 but 6, less than 25% of the participants gave the answer her objection predicts. Thus, it does not seem to be the case that when participants agreed with both sentences for a given pair in Studies 1 and 2, they were treating the objects to be classified as borderline cases. If this is the case, then, the single–concept explanation proposed by our critic fails[21].

5.4. Redefining Conceptual Parthood?

As a final reply, proponents of the hybrid theories of concepts might reject the explanation of the notion of conceptual parthood we proposed in Section 2. Particularly, to accommodate our data, they could reject the idea that the parts of a given hybrid concept should be so organized as to avoid irreducible conflicting judgments. For instance, they could propose that the parts of a given hybrid concept are independent from one another (as the different coreferential concepts are supposed to be, according to the Heterogeneity Hypothesis) and can thus give rise to conflicting judgments.

The trouble with this reply is that what distinguishes hybrid theories from the Heterogeneity Hypothesis is now unclear. If proponents of

21. However, one could defend the Single–Concept Explanation by proposing that the concept expressed, e.g., by "fish" is subject to contextual adaptation. For instance, the weights of the features that constitute a prototype of fish might vary across contexts. We acknowledge that this view of concepts seems consistent with our findings. Further work is thus called for.

hybrid theories were to insist that these two approaches really differ, then they would owe us an account of how they differ. If they were to acknowledge that the differences between the two approaches are now terminological, we would have no more qualms with these revised hybrid theories of concepts. Whether we would then call the different coreferential bodies of information about *x* distinct concepts of *x* or distinct parts of a single hybrid concept of *x* would be immaterial provided it were clearly acknowledged that, far from being coordinated, these bodies are independent.

Conclusion

The explanatory challenge has severely undermined the natural kind assumption: It is not the case that there are many properties common to all concepts. It is however still unclear what approach to concepts is best suited to replace this assumption. In this article, we have compared two approaches, the Heterogeneity Hypothesis and the hybrid theories of concepts. After having characterized both approaches, we have provided some new evidence that *tentatively* supports the Heterogeneity Hypothesis. Our findings suggest that for many speakers, predicates such as "fish", "alive", and "vegetable" express several distinct concepts. At least when they are primed to do so and when a hedge such as "in a sense" is used, people can switch from one interpretation of these predicates to another one — a phenomenon revealed by the willingness of a substantial proportion of speakers to assent to apparent contradictions. This constitutes tentative evidence for the Heterogeneity Hypothesis against the hybrid theories of concepts because, as we cashed out the notion of conceptual parthood, hybrid theories of concepts are committed to the claim that the hypothesized parts of the hybrid concepts cannot result in conflicting judgments.

References

Anderson, J.R., & Betz, J. (2001). A hybrid model of categorization. *Psychonomic Bulletin and Review*, *8*, 629–647.

Danks, D. (2007). Theory unification and graphical models in human categorization. In A. Gopnik and L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 173–189). New York: Oxford University Press.

Gelman, R. (2004). Cognitive development. In H. Pashler and D.L. Medin (Eds.), *Stevens' handbook of experimental psychology*. Vol. 3, *Memory and cognitive processes* (pp. 533–560). New York: Wiley.

Hampton, J.A. (1998). Similarity–based categorization and fuzzyness of natural categories. *Cognition, 65*, 137–165.

Keil, F.C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Keil, F.C., Carter Smith, W., Simons, D.J., & Levin, D.T. (1998). Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition, 65*, 103–135.

Machery, E. (2005). Concepts are not a natural kind. *Philosophy of Science, 72*, 444–467.

Machery, E. (2006). How to split concepts. Reply to Piccinini and Scott. *Philosophy of Science, 73*, 410–418.

Machery, E. (2009). *Doing without Concepts*. New York: Oxford University Press.

Medin, D.L., Lynch, E.B., & Solomon, K.O. (2000). Are there kinds of concepts? *Annual Review of Psychology, 51*, 121–147.

Murphy, G.L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Nosofsky, R.M., Palmeri, T.J., & McKinley, S.C. (1994). Rule–plus–exception model of classification learning. *Psychological Review, 101*, 266–300.

Osherson, D.N., & Smith, E.E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition, 9*, 35–58.

Piccinini, G., & Scott, S. (2006). Splitting concepts. *Philosophy of Science, 73*, 390–409.

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

Rips, L.J., Shoben, E.J., & Smith, E.E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior, 12*, 1–20.

Weiskopf, D. (2009). The plurality of concepts. *Synthese, 169*, 145–173.

Edouard Machery
University of Pittsburgh
machery@pitt.edu

Selja Seppälä
University of Geneva
selja.seppala@eti.unige.ch