Compositionality and Concepts

James A. Hampton

Abstract In this chapter I aim to explain how psychology understands concepts, and why there is a need for semantic theory to take on the challenge of psychological data. All of the contributors to this volume are (presumably) in the business of trying to understand and explain how language has meaning, and the primary source of evidence for this has to be our intuitions of what things mean. Furthermore, if my semantic intuitions (as a theorist) are out of kilter with those of the common language user, then it is my theory which should be called into question and not the lay intuition. This chapter describes a range of results from my research program over the last 30 years, some old and some new, with the aim of giving a general account of using Prototype Theory as a way to explain semantic intuitions.

1 Concepts and Prototypes

In *The Compositionality Papers,* Fodor and Lepore (2002) return frequently to a "knock-down" argument against the suggestion that concepts might be prototypes. Concepts, they argue, must be compositional. It must be possible, if one accepts the representational theory of mind, to explain how the meaning of a complex phrase is based solely on the meaning of the elements from which it is constructed, plus the syntactic structure into which they are placed. To account for our ability to understand the meaning of sentences such as (1) and (2) and countless other similar sentences the semantic system needs a set of fixed symbols to represent the conceptual atoms in the sentences (John, Mary, Bill, loves, hates) which can then be inserted into suitable syntactically structured sentence frames to yield the appropriate meaning for the sentence as a whole.

J.A. Hampton (✉)
Department of Psychology, City, University of London, London, UK
e-mail: j.a.hampton@city.ac.uk

(1) John loves Mary.
(2) John loves Mary, but Mary loves Bill and Mary hates John.

They claim that without this type of compositionality it is not possible to provide an account of how thoughts (and indeed utterances) can express ideas.

Given the claim that concepts must be compositional, the argument continues that prototypes are in fact anything but compositional. For example, the prototypical pet fish is not simply the prototype pet conjoined with the prototype fish. Indeed something like a goldfish or guppy, while being a good match for the pet fish prototype, has little in common with the cats and dogs that are typical pets, or the cod and trout that are typical fish (Osherson and Smith 1981).

More generally, the prototype of any particular complex noun phrase (should it have one) will not be derivable from the prototypes of the content words of which it is composed. Hence concepts (which must be compositional) cannot be prototypes (which are often non-compositional).

The debate on compositionality has generated a large literature in semantics and philosophy—see for example the substantial collection of papers edited by Werning et al. (2012) For critical accounts on whether language is in fact compositional, particularly in respect of the systematicity of its grammar, see Johnson (2004) and Pullum and Scholz (2007). As Pelletier (2017) remarks, the issue probably represents one of the most substantial points of disagreement within cognitive science between the more empirically minded psychologists and linguists who study concepts and word meanings "in the mind", language in use, and lay people's semantic intuitions, and the more theoretically minded semanticists and philosophers whose interests cover broader issues of concepts as constituting lexical meanings within a particular language, and issues of sentential truth and logic. (Like Pelletier, I am speaking generically here).

As an empirical researcher into word meanings, I find myself strongly drawn to the conclusion that lexical meanings do not compose according to the rules of set logic applied to extensions. That is to say that the meaning of a complex phrase will not always be determined simply by the meanings of its components and their mode of combination. I will argue that the construction of complex concepts proceeds (most naturally) through the interactive combination of the intensional meanings of the individual concepts. Wherever two concepts are combined whose intensional contents overlap or interact semantically, then extensional compositionality of meaning will tend to fail. Moreover, if we consider thought rather than language, our capacities to combine concepts greatly exceed the simple combinatorial rules provided by extensional logic, as will become apparent in the last part of this chapter.

1.1 Combining Prototypes

A key piece of evidence for proposing that the category membership of complex concepts depends on more than their components and their mode of combination comes from a series of experiments that I and others conducted in the 1980s and 1990s. To illustrate, in Hampton (1988b, Experiments 2 and 3) I looked at how people interpret the phrases "Sports which are Games", or "Games which are Sports". These phrases were chosen because (at least at first sight) a standard semantic analysis would see the meaning as being a conjunction—that is the extension of both phrases should correspond quite simply to the intersection of the set of Games and the set of Sports. The two alternative phrasings come at the intersection in different ways (either finding the subset of sports that are also games, or vice versa) but the result should be the same.

To test this standard proposal, I provided 36 respondents with a list of 43 recreational activities selected to fall in all four possible combinations of being Sports or not, and being Games or not. Their first task was to judge whether each item was a Sport (together with a rating of degree of typicality or relatedness) and then to judge whether each item was a Game. (Order of tasks was balanced over subjects). Four weeks later the same individuals returned and decided which items were in one or other of the conjunctions mentioned above.

Based purely on their true/false judgments of "whether the general category name can be applied to a particular example", the data were analysed to see whether in fact people followed an intersective rule, only saying Yes to the conjunction for those items (like Tennis or Football) where they had said Yes to both conjuncts. The results were very clear. People did not follow this rule. In fact 25% of games that were not sports, and 54% of sports that were not games, according to first week responses, were nevertheless categorized in the conjunction on week 4. This overextension was not attributable to a contrast effect between the categories, nor to randomness in people's choices, since the equivalent inconsistent pattern of saying No to an item that had been judged as in both sets was much less frequent (11%). A further study showed that the effect was not driven by a response bias, since adding in new items to the list for the second phase so that the expected rate of Yes responses for the conjunction would be 50% (rather than 25%) had no influence on the rate of overextension of the other items.

A final experiment in this paper showed that the effect generalized to six other pairs of categories, including pets which are birds and dwellings which are buildings. Regression analysis was applied to predicting the mean degree of membership in the conjunction from means for each conjunct. Including an interaction term, on average 93% of the variance in mean membership ratings for the conjunction was explained. The analyses also showed that the relative clause form is non-commutative (for example, weapons which are tools are not equivalent to tools which are weapons) and that one of the two concepts often carries more weight in the prediction than the other (a phenomenon termed dominance).

Subsequent research has shown that this lack of respect for the conjunction rule is easily replicated across a range of domains and linguistic ways of expressing a conjunction (Storms et al. 1998; Hampton 1996; Jönsson 2015). It can be shown with the same individual making all three judgments, or with different individuals making each judgment. It also appears when the relative clause is negated. For example, a horse is not often considered to be a vehicle, but it is frequently considered to be a vehicle that is not a machine. Hampton (1988a) showed that the problem of non-extensional combination applies equally to disjunctions. A mushroom is never classed as a fruit, and only 50% of respondents called it a vegetable. However 90% decided that it was either a fruit or a vegetable.

The challenge of these results to extensional theories of semantics should be clear. These results do not just depend on judgments of typicality or even judgments of degrees of truth which might be subject to psychological biases (Osherson and Smith 1981, 1982). They simply reflect the common semantic intuitions of everyday language speakers about the applicability of complex phrases, and as such they require an explanation.

1.2   Intensional Composition

The account of overextension that I offered in Hampton (1987) was a relatively straightforward explanation based on the intensional properties of the concepts involved.

Classically speaking, when a complex concept is constructed as a conjunction, then the features that define it intensionally will be derived in a compositional fashion as a disjunction of the features of the two conjuncts. For example, to form the conjunction of people who are both singers and songwriters, one would look to set up a categorization procedure that would require potential candidates to have the union of the features of each set (person AND singer AND songwriter). To model the combination of categories such as Sports and Games, one can simply propose that Sports which are also Games should therefore be a composite prototype resulting from aggregating all the features commonly associated with either of the two conjuncts. As Sports they should be activities which involve exercise and training and as Games they should also involve competition and fun.

Most importantly, this way of modelling the construction of a conjunctive concept provides a neat explanation of why people are inconsistent in judging extensions of conjunctions. Consider the case where there are two prototype concepts A and B, each with 3 features. In line with Rosch's idea of family resemblance and prototypes, let us suppose that any item will belong in either concept if it has at least 2 of the 3 features of that concept.

Table 1 shows the resulting composite, in which the conjunction A^B is created with all six features in its prototype. Now if having 2 out of 3 features is sufficient to belong in either of the two concepts A or B, then it is natural to suggest that 4 out of 6 features should be sufficient to belong in their conjunction. Accordingly, Item 1

Table 1 A scheme for aggregating the attributes of two concepts A and B into a conjunctive concept A^B. If a "two out of three" rule is used to determine category membership, then Item 1 will be a member of A, B and of the conjunction. However Item 2 will be a member of A and of the conjunction, but will not be a member of category B, having only one of the b attributes. The model predicts overextension of the conjunction through compensation (an excellent member of A can be in the conjunction A^B although only weakly connected to B)

Concept A
a1
a2
a3

Conjunction A^B	Item 1	Item 2
a1	a1	a1
a2	a2	a2
a3	-	a3
b1	b1	b1
b2	b2	-
b3	-	-

Concept B
b1
b2
b3

which has 2 of each of the concept's features should belong in the conjunction. However Item 2, which has 3 A features but only 1 B feature, passes the criterion for belonging in the conjunction, but fails to have enough features to be an example of Concept B. The result is that Item 2 would be overextended—it would belong in the conjunction but not in one of the conjuncts. Effectively, pooling together the features, even before any interaction between them has been considered (resulting in inheritance failure or the addition of new emergent features, see below), leads to the likelihood of inconsistent responding and overextension.

Hampton (1987) collected data on the attributes associated with each of the conjuncts and their conjunction for the same pairs of categories as were used in Hampton (1988b) described above, and traced the way in which attributes for the composite prototype are derived from the constituent components. The procedure involved two separate samples of participants. A first sample was divided into four groups. For a pair of concepts such as Sports and Games, group 1 listed attributes for Sports, group 2 for Games, group 3 for "Sports that are also Games" and group 4 for "Games that are also Sports". They were asked to imagine that they had to define and describe the objects named to someone who was unfamiliar with them. They were to do this by listing on 10 blank lines the attributes or properties that were in any way involved in deciding if an object belongs in the named set. (Full details can be found in Hampton 1987, p. 58). Lists of around 30 attributes for each pair of concepts were then drawn up by including any attribute generated by at least 3 out of 10 participants in any of the four groups.

A second sample of participants was then used to judge the attributes for how important they were for defining each concept and each conjunction. The sample was divided into four groups exactly as before, but now each participant saw the list of attributes and made a rating judgment. "N" was to be chosen as a response if an

attribute was necessary for a category, responses "A", "B" and "C" were used to indicate decreasing importance (A = very important, C = typically true but not very defining), "X" meant the attribute was not usually true, and "XX" meant the attribute was necessarily false of all possible examples of the concept. The results from this rating task were then used to assess how attributes of conjuncts are inherited by their conjunctions.

There were many interesting aspects to the results which I will not try to summarize here. As predicted, importance for a conjunct could be used to predict importance for a conjunction (Multiple R averaged around 0.8, close to the reliability of the scales at 0.85). The process is analogous to one of inherited traits, with traits possessed by both parent concepts being carried through to the offspring conjunction. Most notably there were some attributes which failed to be "inherited" by the composite, and others that "emerged" in the composite which were not in the constituents. Pets which are Birds, for example, lost some features of pets (cuddly) and some features of birds (fly south in winter), but gained other new features not seen in either concept alone such as (lives in cages) or (talks).

1.3 Prototypes as Intensions

I have argued that the way in which people interpret simple semantic rules such as relative clause modification can be accurately modelled by a deeper analysis of the intensional meaning of the words. What is the evidence then that such intensions have the prototype structure that leads to the patterns of overextension seen above?

If we consider most common content words in natural language—nouns, verbs, adjectives and adverbs—then it is often the case that neither the extension nor the intension are easy to pin down. The meaning of function words like prepositions is even harder. Consider for example the following common uses of "on" in English:

(3) The cup is on the table.
(4) I got paint on my shirt.
(5) Harry is on holiday.
(6) The train is on platform 2.

Any attempt to define the extension of situations to which "on" applies is likely to end up simply as a disjunctive list of different cases. The fact that prepositions do not easily translate between languages supports this claim (Bowerman and Choi 2003a, b). When I told a French friend that I had travelled to Paris "sur le train", the puzzled response was to ask if it wasn't very windy sitting on the roof. Similar problems arise with adjectives such as "fresh" or "open" (Murphy and Andrew 1993), with multiple inter-related senses determined by context. For further discussion see Rice (1992).

Returning to the (perhaps) simpler case of nouns, consider a simple everyday term such as "fish". First there is a potential ambiguity arising from the domain of discourse. Fish features in cookery and food and as such its extension may include creatures such as squid, oysters and lobsters. Fish is also part of a commercial

industry, exploiting marine resources, and in the past the category extension included whales. Finally, fish may be taken to have a biological meaning. Unfortunately, those who pin their hopes on science to identify the "correct" extensional class, are due to be disappointed. Current scientific theory suggests there is no common ancestor to the different classes that we call fish. The term describes a disjunctive category with no role to play in biological theory. Thus, not only do we have these three different ways to place the term in context, but within each context the determination of what is in the extension becomes equally problematic. Should shell-fish be included in the culinary term—what about seahorses, rays or squid? The term is underspecified, as the psychological data clearly show (Hampton 1998; Hampton et al. 2006; McCloskey and Glucksberg 1978).

Since intensions are closely tied to extensions, it is not surprising to find that the intensional definitions of terms are equally difficult to pin down. Most encyclopaedia entries will state that all fish are cold-blooded. However, they then go on to say that some fish (like tuna) are not. It has proved very difficult to provide clear definitions for most of our vocabulary terms. There always seem to be exceptions.

Exceptions to the rule that there are always exceptions may be found when a concept has a particularly important role to play in the regulation of society. Then it will often be found to have an explicit definition. The definition of a "US dollar" or "British citizen" has a legal foundation which leads to a clear-cut differentiation into members and non-members of the class.

A slightly less clear example is provided by certain kinship terms, like "father" or "nephew" in English which are often found to be amenable to an analysis in terms of semantic components. As Goodenough (1965) describes it

> A system of kin relationships rests on the established institutions and customs relating to membership in households, sexual rights, the definition of procreation, the legitimization of progeny as members of a jural community and the like.

In relation to "grandmother" Landau (1982) showed how both a definitional criterion (female having a grandchild) and a stereotypical age and appearance are seen in responses for both children and adults when selecting appropriate pictures.

Even when kinship terms such as "uncle" are extended to non-blood relations we are able to distinguish a "real" uncle from other kinds of uncle. On the other hand, the development of non-traditional families has led to the undermining of many kinship terms (see Lakoff's 1987, discussion of "mother"), and terms such as "brothers" and "sisters" can be used with extended meaning to refer to others who share the speaker's beliefs, goals or group membership.

Terms describing crimes (such as murder, theft, or fraud) are likewise provided with definitions by the legislature of each jurisdiction, so that juries can focus on establishing the facts of a case based on the evidence, rather than having to decide how to interpret the meaning of the words. (The latter task is left to judges in higher courts who aim to establish stable interpretations of the terms through reasoned argument about test cases and guessing the plausible intention of the law-makers). However, when a concept does not have this consequential weight resting upon it, it will usually resist easy definition.

The lack of a clear definition where one is needed can lead to expensive court cases, as the following extract from an article by Caroline Davies in the UK newspaper *The Guardian* of 27 April, 2015, shows:

> Bridge, the genteel and physically unchallenging card game played by millions, may exercise the brain muscle, but is it a sport? That is the question taxing legal minds as a high court ruling on Monday paved the way for a courtroom battle to decide. The row centres on a refusal by Sport England to recognise the trick-taking game as a legitimate sport and thus eligible for lottery grants. The English Bridge Union claim it ought to be recognised as a "mind sport" and want Sport England's refusal to do so declared unlawful.

Arguments presented to the court included the amount of physical activity involved (compared for example to rifle shooting), the health benefits of taking part, and the fact that other physical activities are not classified as sports. It was clear that lawyers on each side were seeking to find a plausible definition (intension) that would enable them to either include or exclude bridge from the category containing clear examples of sport such as tennis or football. (What is less clear is why the judge was willing to entertain the argument that the brain is a muscle!)

The issues involved here can be related to two fundamental issues in semantics —context sensitivity and vagueness. Perhaps the lack of clearly specified meanings of terms, and the consequent inconsistency in semantic intuitions results from the lack of a clear context. Alternatively, the difficulty of providing clear meanings may in fact result from those meanings being inherently unclear or vague, in the same way that scalar adjectives such as "tall" or "bald" have been shown to lack precision.

1.4 Context Sensitivity

The meanings of terms can change depending on context. Classic examples are scalar adjectives—a large ant is not as big as a small elephant. Alternatively, there is the example of fish described above, and the different contexts in which the term might be used.

But can context sensitivity fully explain the difficulty in defining extensions and intensions? In an attempt to find evidence for this suggestion, Hampton et al. (2006) ran a set of studies in which we manipulated the context in which people had to classify items in vague categories. We used eight different categories from different ontological domains, and created lists of 24 possible members in which we deliberately included clear cases, clear non-members of the category and about 12 cases that would be difficult to categorise. (The existence of borderline cases in natural categories was originally demonstrated by McCloskey and Glucksberg 1978, when they showed that not only were there many items showing substantial disagreement between people, but that people were also inconsistent in how they categorized those same items when returning a month later to do the task again.)

Our hypotheses for the study related to the idea that apparent vagueness in the category boundaries can be attributed to a lack of a clear context for the classification. We therefore had three main conditions in which we provided different contexts for the categorization. The first was a Neutral control condition which simply asked "Consider each of the following items and decide whether they belong in the category of _____". A second condition, the Pragmatic condition, asked people to categorize items in categories "where people would expect to find them, so that they could be easily found". Scenarios included an internet news group, a mail-order catalogue and a library index. Here we hoped to reduce vagueness because everyone would be attempting to mirror the behaviour of everyone else in the group. The third condition, the Technical condition, provided a set of contexts much like the case of Bridge and Sport described above. People were asked to imagine that they were advising a government agency controlling tax regimes (for Tools and Furniture), ecological reports (for Insects and Fish) or funding agencies (for Science and Sport). They were told that the classification would have important consequences and so they should try to classify "correctly".

Participants worked through each list classifying items as Yes or No, and returned after 3–4 weeks to do the task again in the same condition as before.

Our prediction was that if lack of categorization context was contributing to vagueness, then various measures of vagueness would be reduced in the Technical and Pragmatic conditions. There should be better inter-subject agreement on classification, more stability in categorization decisions over time, a reduced correlation of categorization probability with simple ratings of Typicality, and a shift in the size of the categories, with Technical conditions yielding smaller categories. In the event, none of these predictions was generally supported by the data. Effectively categorization probability in all conditions was correlated at around 0.95 (the limit of measurement reliability) with judgments of item typicality in the category. A second study showed that requiring people to read the instructions aloud and reflect on them before starting to categorize had no effect on the results. There was no easily accessible "deeper" meaning for people to describe if asked to take the task more seriously. Finally, we looked at whether people would be less likely to give a "partial" or graded response in the Technical condition. In this last study, people were given a graded categorization scale to use running from *not at all* through *barely, sort of* and *very much,* to *completely*. If people felt that a category has a clear definition (even if they are uncertain what it is), then we expected them to be disinclined to use a partial rating such as "sort of" or "very much", and to stick to the two extreme responses—"not at all" or "completely". In the event our manipulation of context had no effect on this measure either.

In the light of these results, it would appear that instability and disagreement in categorization is not exclusively driven by a lack of specificity in the context. Whether simply classifying, trying to capture common categorization practices, or yet advising a technical committee on the correct way to classify, people rely on the same underlying conceptual representation and this is best described in terms of a typicality gradient.

Typicality is a measure that has had wide use in Psychology, but its relation to semantics is often rather obscure, and subject to misunderstanding. In the next section I therefore discuss in detail just what typicality is measuring, and how it relates to issues of semantics.

1.5   Typicality and Gradedness

In Hampton (2007) I define a position on the relation of typicality to graded categorization. In the Threshold Model I propose that a semantic category is represented by a set of intensional information in the form of a prototype (which may include schematic structure about causal-explanatory links between features). Potential exemplars can be ranked in terms of their similarity to this prototype, as determined by an asymmetric measure of how well the exemplar matches the prototype features. For example, similarity of a tomato to fruit will be greater on this measure than the similarity of fruit to tomato. It is assumed that tomato matches more features of fruit than fruit matches features of tomato because of the greater abstraction of fruit.

The ranking based on similarity then provides the basis for judgments of typicality (assuming other factors are held constant—see below), and also provides the basis for categorization decisions through the application of a criterion or threshold. With the additional assumption that the placement of the threshold is subject to normally distributed error both within and across individuals, a standard psychometric function is obtained relating the probability of a positive categorization to the underlying similarity.

It is perhaps common in some parts of the Cognitive Science community (e.g. Armstrong et al. 1983; Fodor 1998; Osherson and Smith 1981, 1982, 1997) to dismiss typicality effects as purely psychological and hence peripheral to the development of lexical semantic theory. After all, we know that the way in which the mind stores words in the mental lexicon shows all kinds of psychological influences that are orthogonal to issues of lexical meaning. Frequency of a word in the language, for example, has large effects on reading, memory and a range of other cognitive tasks. It has also been found that the degree to which words are associated (like "Fish" and "Chips" in the UK) can be highly predictive of a range of phenomena. Why should not typicality effects be of the same kind?

I will argue that when people judge typicality (for example of an item in a category) they may be judging quite a number of different things. Nonetheless paramount among those different dimensions is similarity to a prototype representing the common intensional properties of the class. Because this measure of similarity also determines the degree to which an item can be said to belong in a category, typicality ratings do an excellent job of predicting the likelihood that an item will be placed in a category as evidenced in the Hampton et al. (2006) experiments, and many other similar studies.

1.6   Does Variation in Typicality Really Undermine the Classical Model?

Any description of the impact of Rosch's prototype theory on the psychology of concepts tends to make much of typicality effects. The classical theory against which Rosch was arguing proposed that a concept could be defined intensionally as a conjunction of individually necessary and jointly sufficient features. A concept such as "bachelor" could be defined as "human, male, adult, and eligible to marry". It has been claimed by supporters of the prototype model that this classical theory gives equal status to all items that meet the definition, so there should be no differences between items in terms of how well they represent the class. Although often repeated, I do not take this to be a fair criticism of the classical model. Typicality can reflect many different underlying structural variables, many of which do not relate to the question of whether a concept term applies to an exemplar. It is important therefore to tease apart the different influences on typicality to get a clear picture of the role that it plays in conceptual structure.

First, the notion of typicality or goodness-of-example is often confounded with other non-semantic dimensions such as familiarity (Malt and Smith 1982). Indeed Armstrong et al. (1983) demonstrated that well-defined categories such as Odd and Even Numbers have clear typicality structure, most probably based on simplicity and familiarity (but see Larochelle et al. 2000, for counter-evidence). (Of course, the demonstration that well-defined categories show typicality effects does nothing to undermine the theory that for other types of categories, lacking an explicit definition, typicality may be a critical factor in determining membership.)

Second, it has been suggested that category membership is determined by a defining core of features, whereas additional "characteristic" features are associated with typicality differences (Osherson and Smith 1981; Rey 1983; Smith et al. 1974). In support of this proposal, Rips (1989) presented a variety of attempts to dissociate measures of similarity, typicality and category membership, in which it is claimed that some item may be more typical of category A than of category B, even though it is a better member of B than of A. His results have however not stood up well under replication (Hampton et al. 2007; Smith and Sloman 1994).

Given that Typicality effects on their own do not provide strong evidence against the classical model, what do we know about their basis?

1.7   Ingredients of Typicality 1: Ideals

It has been shown (e.g. Barsalou 1985) that when people are asked to say how typical an item is as a member of a category, then they are influenced by several different dimensions. First and foremost, Typicality ratings are assumed to be a pure measure of the underlying similarity in meaning, or degree of match of semantic features, between a member and its superordinate category. This dimension of

similarity is clearly a major influence on Typicality ratings. However other factors are also involved. Barsalou (1985) showed that in addition to what he termed Central Tendency (closeness of a concept to the centre of its category), ratings of Typicality were also correlated with frequency of instantiation (a measure of familiarity) and matching of Ideals. An ideal is a feature of a concept that represents extreme rather than average values of a dimension. Thus a winter coat may be considered most typical if it is ideally warm and light, as opposed to being closest to the average winter coat (which will be of average warmth and average weight).

So Typicality per se must be interpreted with this ambiguity in mind. The ambiguity has been made a lot worse by Rosch's (1975) original characterization of typicality as "goodness-of-example", a term also used in Barsalou (1985). Barsalou notes that the word "typicality" was not used in his study because it could bias participants into thinking of frequency of instantiation. As a consequence, he asked for goodness-of-example, with a scale running from "poor example" to "excellent example". Similarly, Burnett et al. (2005) concluded that expert fishermen judged the typicality of types of fish based on ideals, while using "goodness of example" as their measure of typicality. The difficulty here is that asking about "goodness" leads to an evaluative judgment and hence allows ideals to have a greater influence on the judgments.

There have been very few attempts to distinguish between the two senses of typicality. One exception is Kittur et al. (2006) who dissociated the two dimensions using a novel relational concept learned in the laboratory. They found that ratings of goodness of example reflected just ideals, whereas typicality judgments reflected both ideals and central tendencies.

One way to understand the relation of typicality and ideals would be to propose that ideals should be understood as contributing to typicality itself which then determines degrees of membership. In this way, typicality would mediate the influence of ideals on category membership. This proposal needs empirical testing. In an experimental manipulation of ideals, Kim and Murphy (2011) demonstrated that in fact ideal exemplars that best served a category's goals were not necessarily perceived as most typical. For example, a great party might be considered ideal, but was not judged as typical.

1.8   Ingredients of Typicality 2: Frequency and Familiarity

As well as Ideals, Barsalou (1985) also identified Frequency of Instantiation as a component of Typicality. Participants were asked to judge subjectively how often a category member occurred as an instantiation of the category. Allied to this measure is a second measure: Category Dominance. Going back to the early days of associationist psychology, the measure of Category Dominance is the relative frequency with which an item is generated when people are asked to list all the category members that come to mind within a limited time (Battig and Montague 1969). A third measure related to frequency of instantiation is familiarity, in which

participants rate items in terms of how familiar they seem. These different measures tend to correlate together and to form a separate dimension from typicality owing to family resemblance or similarity (Barsalou 1985; Hampton and Gardiner 1983).

Hampton (1997a) was able to show a double dissociation of the effects of Typicality and Category Dominance on response times and errors when people make speeded categorization decisions. In a first experiment, regression methods were used to differentiate the effects of typicality, familiarity and category dominance on the average time to categorize items, and the likelihood of making a positive response. Participants were given a category (e.g. Fruit) and then a list of words one at a time. They had to make a speeded decision for each word whether it belonged in the category or not. Mean reaction time was predicted in a multiple regression equation using norms for semantic categories collected by Hampton and Gardiner (1983). (Hampton and Gardiner 1983, used instructions for typicality that explicitly differentiated it from frequency of instantiation.) Likelihood of a Yes response was also predicted. The results showed that typicality and category dominance each made independent and significant contributions to predicting decision time. Although the two measures were correlated with each other, the speed in making a decision was driven both by the availability of the item in memory (as measured by category dominance) and by the similarity of the item to the category (as measured by typicality). When it came to response probability, only typicality predicted the likelihood of a Yes or No response.

The second experiment introduced two manipulations. First, manipulating the difficulty of the task by including closely related false items (e.g. a bat is a bird) slowed down atypical items relative to typical items, but had no effect on items as a function of their category dominance. A second manipulation in which half the items were seen in a different context before having to be categorized showed that the category dominance effect but not the typicality effect was eliminated by earlier exposure of the items. Taken together the results all suggest that while high dominance items are more readily available in memory, the actual decision of whether something is in a category is just affected by its typicality and not by associative strength. There is an interesting parallel here with the heuristics of Availability and Representativeness proposed by Tversky and Kahneman to explain people's judgments of probability (Kahneman et al. 1982).

1.9 Typicality and Membership

Having described the multiple influences on typicality, including the ambiguity of what it is to be a "good" example, and the confounding with familiarity and category dominance, what is the evidence that there is nonetheless a purer notion of typicality that should be taken seriously as a component of meaning? The results of the study by Hampton et al. (2006) described above provide one such piece of evidence. Here we had eight categories in which there was uncertainty or vagueness in the classification of borderline items. Moreover, the basis of the disagreement

and inconsistency did not come down to mere ignorance. There is no fact of the matter or correct answer to these borderline categorization cases. In that respect they can be termed an example of Vagueness, with similar properties to the traditional cases of vagueness seen in adjectives like red, bald or tall. Being bald is a matter of degree, so that it is meaningless to ask exactly how many hairs need to be lost before someone is correctly termed bald. In the same way, many noun categories have membership which is a matter of degree (sociology may be considered a better science than palm-reading) but there is no hard and fast way to determine who is right and who is wrong in the event of a borderline dispute. We come back again to the dispute about bridge being a sport. With due respect to the UK Supreme Court, there is no higher authority to which one can turn to decide the question in an objective fashion (as there might be in the case of a biological or technical term). As in matters of taste, it appears that each may be entitled to his/her own opinion about such cases (Wright 1995).

It is in the context of this vagueness in noun meanings that the notion of Typicality can be helpful. Disputes about borderline cases often end up with party A arguing "X is a sport because it has features D, E and F", while party B argues "X is not a sport because it lacks features P, Q and R". But this is exactly what the "pure" notion of Typicality captures—the fact that the more features of a concept an item possesses the more justified one is in placing it in the category. There is a continuity between one category member being more typical of a category because it has more matching features (as in the case of a robin being a more typical bird than an ostrich, even though both are clearly birds) and one item being more likely to be classed as a member of a category than another for the same reason (as in ten-pin bowling being considered more of a sport than billiards.)

This is a critical point for the debate between formal semantics and psychology. If typicality is a purely psychological phenomenon that does not affect truth values (as in the robin versus ostrich case) then it is safe for semantics to ignore it and instead to focus on category membership (a position taken among others by Osherson and Smith 1997). However when other conceptual categories are considered, it turns out that variation in typicality (as determined by similarity to a prototype or by the degree to which something possesses the prototypical features) does affect truth values. In the Hampton et al. (2006) experiments we showed again and again that rated typicality was the best predictor of people's judgments of truth for sentences such as "seaweed is a vegetable", "a tomato is a fruit", "a squid is a fish" or "a piano is a kind of furniture".

In Hampton (2007) I argued therefore for a single underlying dimension that in the first place determines how typical some item is of its class (in the family resemblance sense of typical, rather than anything involving ideals or familiarity), and in the second place determines how much of a member of the class it is. This underlying dimension relates to the degree to which the conceptual representation of the item brought to mind in the given task context matches that of the superordinate category that it is being compared to.

To develop this model, the notion of Typicality has to be extended or adapted a little further. Linguistically speaking, to say that X is a typical (or atypical) Y carries the presupposition that X is indeed a Y. One would probably not say that a bat is an atypical bird. It may resemble a bird but that doesn't make it an atypical one. However in the original introduction of typicality ratings as "goodness-of-example" Rosch (1975) also chose to ignore this refinement and asked participants to judge typicality across a full range of items running from typical exemplars through borderline cases to clear non-examples. Adapting to the pragmatics of the task, her participants duly obliged, and so Typicality also has an extended meaning corresponding to something like "typicality if it is a member and closeness to the category if it is not".

In a number of papers, (Hampton 1979; Hampton and Gardiner 1983; Hampton 1988b) I developed a graded membership scale based on this more explicit notion which was then used in the studies of concept conjunction negation and disjunction described above. As discussed above, the key result of those studies was that membership in a conjunctively defined category showed a continuous gradation (in terms of the probability of a positive response) that was highly predictable in a regression from degree of membership in the two constituents. Furthermore, just as typicality in a conjunction is known to sometimes surpass typicality in a conjunct (the guppy as a pet fish is the example proposed by Osherson and Smith 1981), so **membership** in a conjunction can surpass membership in a conjunct.

This phenomenon of overextension has been demonstrated most recently in a study of activity verbs by Martin Jönsson. Jönsson (2015) showed people videos of an actor simultaneously performing two actions, such as Smoking and Walking. The action would be a typical example of the first action (e.g. smoking) but a very atypical example of the second (e.g. walking). The task asked for a Yes/No answer to simple questions such as "Is this man smoking?", "Is this man walking?" or "Is this man smoking and walking?". Jönsson found with this particular example that 100% answered yes to the first question, and only 39% answered yes to the second. However 70% answered yes to the conjunctive question. It appears that likelihood of agreeing to classification in a conjunction may involve an average of perceived degrees of membership in each conjunct, rather than the likelihood of believing that the item is in the first category and also believing that the item is in the second.

A crucial test of the involvement of Typicality in categorization judgments is to show that variation of typicality among cases which are **clearly members of a category** can nonetheless affect categorization in a conjunction. Suppose that an item is clearly a member of one category, but is on the borderline for another. For example, suppose that everyone agrees that Jack is bald (about half his head is hairless) but only 50% agree that he is tall. Then the composite prototype account would predict that increasing Jack's degree of baldness yet further would compensate for his lack of tallness and make it **more likely** that people would accept that he is both bald and tall.

This prediction of compensation between typicality in one category and membership in a conjunction was tested in Hampton (1996). People made a set of three judgments about cartoon faces representing a range of age from child to adult and a

range of emotion from happy to sad. Critical test items were clearly either children or adults, but differed in typicality in those categories. At the same time these test faces were borderline in terms of emotion. The study showed, for example, that variability in the typicality of a given adult face influenced the likelihood of categorization in a conjunction such as "happy adult". Thus typicality of a face as an adult face could compensate for borderline membership as a happy face and affect the categorization in the conjunction.

1.10   Differentiating Vagueness from Ignorance

A recent paper (Hampton et al. 2012) shows that the ontological uncertainty about whether (say) bridge is a sport can be differentiated from other kinds of uncertainty based on ignorance. In a set of studies we looked at the problem of higher order vagueness. The basic set up was as follows. Two groups of people had to judge whether a list of items belonged in a category, and as previously, the list was designed to contain many borderline or disputable items. The procedure required them to return after two weeks and to perform the task again (as in McCloskey and Glucksberg 1978). We measured the likelihood that they would give the same response on each occasion—a measure we labelled as Consistency. In the first group who responded simply Yes or No, people would typically maintain the same response for about 80% of items. People in the second group, rather than saying simply Yes or No, were given the chance to create a third, middle, response category. They were instructed to first decide if the item was 100% certain to be in the category, or 100% certain to be NOT in the category, and respond accordingly. Any item for which they were not 100% sure, they were told to put in a middle response category of "Not 100% sure". As for the first group, we measured consistency of responding.

Our original intuition was that this second condition would lead to greater consistency. As people could "cherry pick" the easy items and leave the others aside, and as they would be given credit for being consistent if they put an item in the "Not 100% sure" category both times, we felt that they would change their minds much less often. In fact, the results were quite clear in showing that the likelihood of changing your mind about whether something is 100% certain to be in a category is no more nor less than the likelihood of being inconsistent in judging if it is in the category or not. Higher order vagueness (determining the boundaries of the vague region where things are unsure) turned out to be just the same as lower order vagueness (the indecision within the vague region).

This effect can be used to argue that vagueness about category membership is not equivalent to uncertainty owing to ignorance. In different versions of the task, we tried the same procedure with general knowledge statements. Instead of a statement such as "rhubarb is a fruit" we had statements like "The Uruguayan flag has red in it". Now, the second group who were allowed to say when they were unsure were significantly more consistent in their responses. After several other

studies, including both those in the paper and other unpublished studies done since, we can conclude that when there is an objective truth to a statement (as in general knowledge, or as in the correct meaning of a word, or as in memory for a video) then people can reliably identify the statements that they "know they don't know". On the other hand, when the truth of a statement has a more subjective basis, as in categorization but also as in personal judgments about one's aspirations, moral beliefs or early childhood memories, then one cannot do so. In these cases, asking people to only say "yes" when they are definitely sure simply moves the decision criterion to a higher level but does nothing to reduce the inherent unreliability of the decision.

A recent unpublished study, conducted with Shauna-Kaye Williams demonstrates this effect with a single dimensional example of vagueness. We had two sets of faces which varied according to their emotional expression. The first were a set of morphs between a neutral expression and a happy expression. The second were morphed between a neutral expression and a surprised expression. In a first session, participants went through each set of faces twice, once deciding if the faces were happy or not, or surprised or not (depending on the set), and once deciding if they were *clearly* happy or not, or *clearly* surprised or not. They returned for a second session a week later, and repeated the task. In this case, we were interested in whether asking people only to respond positively if the faces were "clearly" showing the emotion would lead to a sharper boundary and more consistent responding over time. In fact, neither of these occurred. Figure 1 shows the results where logistic regression functions were fit to each individual's data and then a plot made based on the average slope and threshold. In both sets of faces, the requirement to select faces that clearly showed the expression simply moved the threshold to the right, while leaving the sharpness of the boundary unaffected. (In fact, the surprised faces had a slightly lower slope—a vaguer boundary—for the clearly judgments).
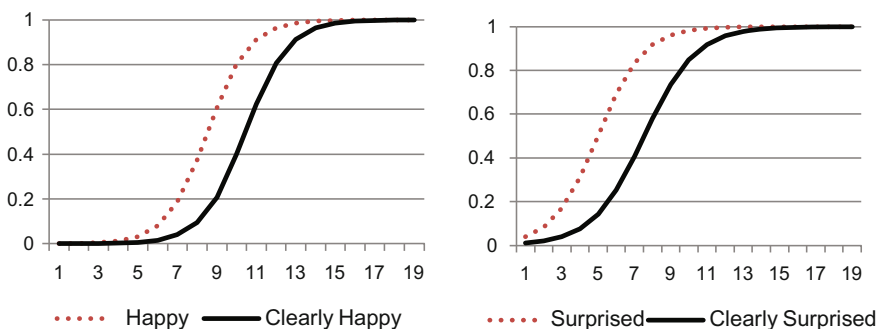

Fig. 1 Logistic functions fit to the likelihood of a positive categorization as a function of the morphed scale of emotional intensity for happy or clearly happy faces (*left*) and surprised or clearly surprised faces (*right*)

1.11   Concept Intensions as Fundamental

I have hoped to show that there are a number of phenomena that argue for a fundamental role for intensional representations of concepts as the raw components of thoughts. My research into conjunctions of prototype concepts shows that people are not particularly concerned to respect the logic that requires that to be in a conjunction means to be in each of the conjuncts. Likewise there is good evidence that neither disjunction nor negation fares any better in terms of maintaining logical norms (Hampton 1988a, 1997b). To account for these deviations from logic, it is most fruitful to look at how intensions might be combined in different ways to form conjunctive, negated or disjunctive concepts (Hampton et al. 2012). The Composite Prototype Model (CPM, Hampton 1987, 1988b) details how we might account for the results through a process of integrating two prototype concepts into a single composite to represent the complex concept. As Pelletier (2017) describes it, this is compositionality as applied to prototypes. We know that prototypes for conjunctions (e.g. pet birds) may look quite different from those of their conjuncts (pets and birds considered separately). As described above, the model suggests an initial process of attempting to combine the two concepts by taking the disjunction of their features. This is followed by the identification of points of incompatibility and either the deletion of certain features (pet birds do not migrate in winter) and/or the addition of new emergent features not seen in either concept (pet birds live in cages though neither birds nor pets do normally).

In order to explore the process of conceptual combination in a "pure" state, without the influence of prior familiarity, Hampton (1997c) described a study in which people were required to describe conjunctions of categories which do not currently exist, such as a computer that was also a kind of teacup, or a vehicle that was also a kind of fish. Why set people such a task? I was primarily interested in demonstrating that concept intensions are flexible and adaptable. Just as when forming the concept of Pet Fish one must abandon some salient features of each concept (we don't eat pet fish with French fries, and nor do they cuddle on our laps), so when truly incompatible conjunctions need to be imagined the process can be taken to extremes. Figure 2 shows an example of one of the solutions offered by

Fig. 2 A bird that is a kitchen utensil. Redrawn from original drawing of an anonymous participant
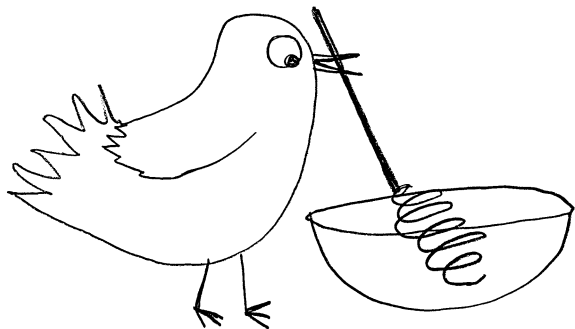
the more creative participants, in response to the requirement to find a bird that was also a kitchen utensil. In this case a woodpecker has been trained to whisk eggs using its powerful head movements.

Qualitative analysis of the results of this first pilot study showed some very interesting principles were involved in combining concepts. First, there was evidence for instantiation of superordinate categories to basic level categories—birds became woodpeckers, fruit bananas, and furniture couches. Second, people would align properties and functions between the two concepts. The example given here shows how the need to find a function for a kitchen utensil is met by finding a behavior of a particular bird that can serve the function of a particular utensil. Third, there was evidence for simulation processes, in line with Barsalou's suggestions about the role of concepts in achieving goals (Barsalou 1991). Commentary provided by the creator of the woodpecker whisk pointed out that it would not need electrical power (good for camping trips) but would on the other hand be unhygienic. The use of simulation takes the specification of the combined concept and develops it in totally non-compositional directions, as the new concept is adapted to real world knowledge. Finally, many solutions identified conflicting properties in the newly combined concept, and offered new emergent features to resolve them. For example, when a participant solved "A fruit that is also a kind of furniture" by proposing a banana couch, they went on to specify that the banana had been modified to grow very large and to ripen very slowly.

Extensional accounts of meaning focus on the sets of exemplars in the world. As such they can have nothing to say about concepts which have no members. (The problem of empty and fictional names is well known, Braun 2005). Thus when we think about counter-factual or hypothetical objects, the extension has to be taken to refer not to the actual world but to a virtual or "possible" world. Computer teacups do not currently exist, but it is possible to imagine an alternative world in which they do, and we can speculate about their properties. In fact this process of conceptual combination may be a key factor in innovation and creativity.

In a recent paper (Gibbert et al. 2012) my colleagues and I investigated creativity in the forming of concepts of hybrid products—artifacts that serve more than one traditional function. In direct agreement with the processes postulated by the Composite Prototype Model we showed how first attempts to imagine (say) a pillow that was also a telephone would simply aggregate the features of the two concepts into a composite. However when the two concepts being combined were sufficiently dissimilar, then a second attempt at combining the concepts would generate integrative solutions in which features of one concept would be aligned with those of the other to provide emergent functionality. The telephone could be programmed to provide gentle sounds to help people get to sleep, or the pillow could gently move to alert the user that a phone call was arriving. A less kind proposal was that the pillow would allow one to nap while listening to one's mother on the phone. When people were able to generate integrative solutions, these were consistently judged to be more likely to succeed as marketable products.

1.12 Impossible Objects and Hierarchical Levels

To conclude this chapter I will describe a study in this project, conducted with Diane Lewis and Zachary Estes in which we looked more specifically at the issue of instantiation. We designed a study in which half the concepts were basic level categories (Rosch et al. 1976) corresponding to the most common names for objects, like CAR and LEMON and the other half were superordinate categories like VEHICLE or FRUIT, where there is no single image than can be formed of the category as a whole.[1] We had two ideas in mind about how this variable would affect the likely success of forming incompatible conjunctions. On the one hand, superordinate categories place fewer constraints on the solution. For an object to be a car requires some minimum description in terms of shape, size and material which need not be true for something to be a vehicle. We therefore predicted that in general our participants would find it easier to combine disjoint sets at the super-ordinate level, leading to Superordinate-Superordinate combinations being the easiest to form, Basic-Basic combinations to be hardest, and mixed combinations in between. As a rider to this hypothesis, we also considered it likely that successful use of Superordinate concepts would be found greatest when they were actually instantiated in the solution as familiar basic level concepts.

Set against this prediction was the possibility that because basic objects are very familiar and easily imagined, there would be some processing advantage to having the task set at this concrete level. Indeed there is evidence (Smits et al. 2002) that people actually make decisions about superordinate categories by retrieving prominent basic level exemplars. In particular, we felt that a case could be made for arguing that the mixed conditions Superordinate-Basic and Basic-Superordinate would be the easiest. In these conditions one of the concepts would be anchored to a familiar and concrete basic level object, and the task would then be to modify or transform this easily imagined object in order to meet the criteria of the other superordinate category in some way.

The design was a 2×2 within subjects design manipulating whether the first and second nouns were basic or superordinate concepts. Thirty-two students and other young adults (18 female) aged 17–32, completed booklets. Eight sets of items were constructed. Each set was a quadruple of a pair of basic level concepts to be combined, together with their superordinates. Table 2 lists the materials used. A combination was created by taking one of the nouns from the left two columns and combining it with one of the nouns from the same row in the right two columns. There were thus four possible combinations for each quadruple, such as Banana-Bus, Banana-Vehicle, Fruit-Bus and Fruit-Vehicle.

Each participant was given two pairs in each of the four conditions, and the materials were rotated through the four conditions across four groups of participants

---

[1]Some of our superordinate concepts, like Bird and Fish were probably at the basic level in terms of Rosch et al.'s (1976) criteria. However the manipulation of level was still valid with "Wood-pecker" providing greater constraints on a combined concept than "Bird".

Table 2 Materials used in the experiment for forming "impossible" combinations of concepts

First noun		Second noun	
Basic	*Superordinate*	*Basic*	*Superordinate*
Banana	Fruit	Bus	Vehicle
Woodpecker	Bird	Jumper	Clothing
Shark	Fish	Daffodil	Flower
Table	Furniture	Mosquito	Insect
Bomb	Weapon	Carrot	Vegetable
Snake	Reptile	Bungalow	Building
Horse	Mammal	Spanner	Tool
Hamster	Pet	Guitar	Musical instrument

Fig. 3 Success of solutions as a function of level of the head and modifier nouns
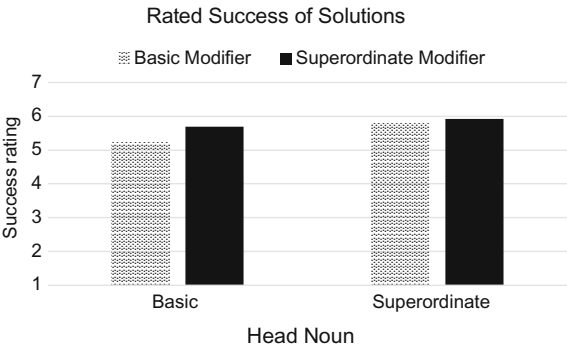
so that each quadruple contributed equally to all conditions. The first nouns were always the Head nouns, and the second the Modifiers in a phrase such as "A banana that is also a bus".

1.13   Results

Responses were rated by two independent judges on scales of success (1–10) and symmetry (rescaled for analysis as 1 = bias towards head, 0 = no bias, −1 = bias towards modifier). Reliabilities were estimated by the Spearman Brown method as 0.70 for success ratings and 0.75 for symmetry. Raters disagreed by more than 4 scale points for the success rating on only 5% of occasions, and gave opposing symmetry judgments also on only 5% of occasions.

Success. Figure 3 shows the mean rated success of solutions as a function of the level of the head and modifier concepts. Problems involving superordinate terms were more successfully solved than those involving basic level terms, both for the head noun (5.8 vs. 5.4) and for the modifier noun (5.8 vs. 5.5). ANOVA confirmed independent significant main effects of level for the head noun ($F(1, 31) = 6.9$,

$p < 0.05$, and for the modifier noun ($F(1, 31) = 4.55$, $p < 0.05$), with no significant interaction ($F < 1$).

Symmetry. Analysis of the Symmetry judgements showed that solutions tended to be more similar to the head noun than the modifier (mean bias = 0.14, SD = 0.54, $t(127) = 2.84$, $p < 0.005$), but there were no effects of level. This result suggests that the head noun may be taken as a starting point and modified in the direction of the modifier, rather than vice versa.

Instantiation. The significant effects of level on success were in keeping with the prediction that the greater flexibility allowed to the participant from the use of a superordinate would allow better solutions to be found. To test this notion further we examined the interaction between success and the amount of instantiation used in a solution.

Solutions for the conditions involving superordinates (Basic-Superordinate, Superordinate-Basic and Superordinate-Superordinate) were divided on the basis of whether either the head or modifier superordinate was clearly instantiated as a particular basic level term. Some of the 16 superordinate categories were almost always instantiated (notably furniture, fruit, mammal, pet, vehicle) whereas others were almost never instantiated (flower, fish). The likelihood of a superordinate concept being instantiated was greater (72%) when the other concept was also superordinate, than when the other was a basic level term (59%), so it appears that finding a solution is easier if **at least one** of the terms is at the basic level or has been instantiated at the basic level.

Average success of solutions was compared for cases where the superordinate(s) were instantiated and cases where they were not. For this purpose instantiation was treated as a Post Hoc Factor. For the conditions with one Basic and one Superordinate concept, instantiation of the superordinate had a relatively small effect (mean success = 6.0 instantiated, 5.4 un-instantiated, $t(126) = 1.97$, $p = 0.051$).

However for the Superordinate-Superordinate condition, instantiation had a sizeable effect on success. Where neither was instantiated (13 out of 64 cases), mean success was only 4.3, whereas when either one or both were instantiated it rose to 6.3. Because of small cell sizes, a one-way ANOVA was conducted comparing three levels of instantiation—both nouns instantiated, just one instantiated, or neither noun instantiated. There was a strong effect of instantiation ($F(2, 61) = 6.1$, $p < 0.005$). Figure 4 illustrates some of the more successful solutions offered.

What has been shown by these explorations of the creative potential in our conceptual system? Clearly the primary function of words and sentences is to enable us to communicate and coordinate our thoughts about the world. Lexical items have a dictionary meaning that provides a firm basis for learning a language and using it effectively across a range of social contexts. At the same time, the concepts that constitute the basis of those meanings are capable of showing a flexibility that is fundamental to the process of invention and conceptual change. To understand how lexical composition occurs it helps to understand more about how conceptual contents can be combined. That process would appear to require access to the full repertoire of human cognitive capacities, well beyond the limits of a set
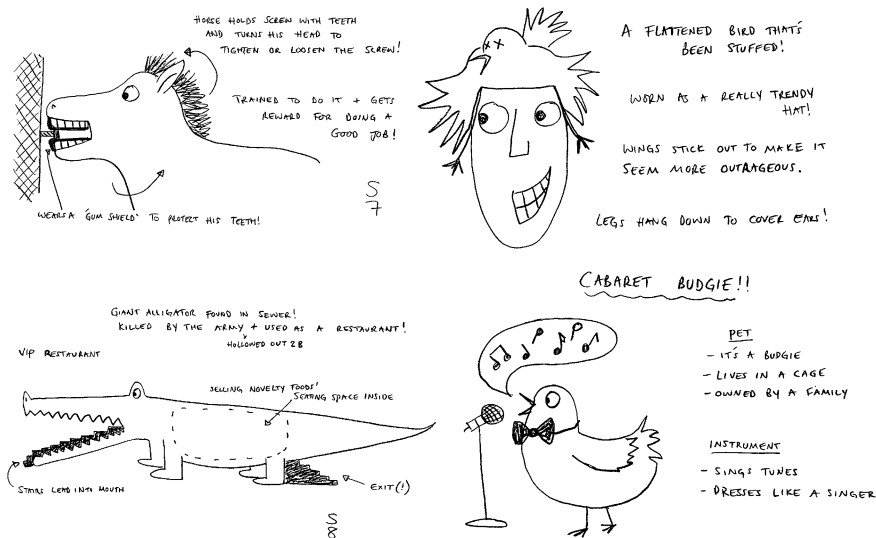
Fig. 4 Examples of the successful solutions generated by the more creative participants. Clockwise from *top left* a horse which is a tool, a bird which is clothing, a pet which is a musical instrument, and a reptile which is a building. Instantiation, alignment and emergent features are evident in these solutions. Original drawings of anonymous participants have been redrawn

of compositional rules applied to a finite set of fixed concepts. In particular it may require active simulation of concepts in an imagined world (Barsalou, 2017) and access to knowledge of the world (Murphy and Medin 1985).

2 Conclusions

In this chapter I have tried to provide some hard evidence that lexical meanings are often concepts that are constituted as prototypes in the mind of the language user. The prototype consists of a set of correlated features that represent what people know and expect about the most typical or representative example of the kind, and the amount of variability that can be expected around that. Because these concepts lack hard definitions, it is common to find borderline disputes about meaning. Furthermore when modelling the way in which people interpret apparently conjunctive phrases such as "an A which is a B" it is necessary to take account of these prototype intensions in order to explain the patterns of overextension and compensation that occur, together with other effects such as non-commutativity and category dominance.

I hope also to have provided an explanation of why typicality should not be taken as a unitary measure, since as a task it can invite a range of different pragmatic interpretations, including ideals and familiarity as well as the intended one of representativeness.

Let me conclude with some comments about the relation of this work to the work of formal semantics as described in the chapter in this volume by Pelletier. It is perfectly true that the work I have described offers no account of the difference between individuals and kinds, no account of how the scope of quantifiers is determined from syntax and no account of indefinitely many other linguistic and semantic phenomena. That was never its aim. The focus on intersective noun combinations was primarily to demonstrate that you need intensions to explain people's intuitions of applicability in these cases. There is a large psychological literature on other forms of conceptual combination involving noun-noun compounds (Wisniewski 1997; Gagne and Shoben 1997; Estes and Glucksberg 2000) showing that intersective interpretations are relatively rare compared to a number of other commonly used thematic relations such as "MADE OF" (e.g. CHOCOLATE EGG), "LOCATED IN" (e.g. CITY BUS) or "USED FOR" (e.g. CEMENT TRUCK). The question is therefore whether the two distinct approaches to semantics can find a way to mesh, or whether there are fundamental incompatibilities between them. A problem of finding the right conceptual combination.

References

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

Barsalou, L. W. (1991). Deriving categories to achieve goals. *The Psychology of Learning and Motivation: Advances in Research and Theory, 27,* 1–64.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 11,* 629–654.

Barsalou, L.W. (2017). Cognitively plausible theories of concept composition. In J. A. Hampton & Y.Winter (Eds.), *Compositionality and concepts in linguistics and psychology* (pp. 9–30). Berlin: Springer.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph, 80*(3, part 2).

Bowerman, M., & Choi, S. (2003a). Space under construction: Language specific spatial categorization in first language acquisition. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and cognition* (pp. 387–428). Cambridge: MIT Press.

Bowerman, M., & Choi, S. (2003b). Space under construction: Language specific spatial categorization in first language acquisition. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought*. Cambridge MA: MIT Press.

Braun, D. (2005). Empty names, fictional names, mythical names. *Noûs, 39,* 596–631.

Burnett, R., Medin, D., Ross, N., & Blok, S. (2005). Ideal is typical. *Canadian Journal of Psychology, 59,* 3–10.

Davies, Caroline (2015). High court ruling paves way for courts to decide whether bridge is a sport. *The Guardian, 27/04/2015.*

Estes, Z., & Glucksberg, S. (2000). Interactive property attribution in concept combination. *Memory and Cognition, 28,* 28–34.

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

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

Gagné, C. L., & Shoben, E. J. (1997). Influence of thematic relations on the comprehension of Modifier-Noun combinations. *Journal of Experimental Psychology: Learning Memory, and Cognition, 23,* 71–87.

Gibbert, M., Hampton, J. A., Estes, Z., & Mazursky, D. (2012). The curious case of the refrigerator-TV: Similarity and hybridization. *Cognitive Science, 36,* 992–1018.

Goodenough, W. H. (1965). Yankee kinship terminology: A problem in componential analysis. *American Anthropologist, New Series, 67*(5), 259–287.

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. (1988a). Disjunction of natural concepts. *Memory & Cognition, 16,* 579–591.

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

Hampton, J. A. (1996). Conjunctions of visually-based categories: Overextension and compensation. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 22,* 378–396.

Hampton, J. A. (1997a). Associative and similarity-based processes in categorization decisions. *Memory & Cognition, 25,* 625–640.

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

Hampton, J. A. (1997c). Emergent attributes in combined concepts. In T. B. Ward, S. M. Smith, & J. Viad (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 83–110). Washington DC: American Psychological Association Press.

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

Hampton, J. A. (2007). Typicality, graded membership and vagueness. *Cognitive Science, 31,* 355–383.

Hampton, J. A., Aina, B., Andersson, J. M., Mirza, H., & Parmar, S. (2012). The Rumsfeld effect: The unknown unknown. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 38,* 340–355.

Hampton, J. A., Dubois, D., & Yeh, W. (2006). The effects of pragmatic context on classification in natural categories. *Memory & Cognition, 34,* 1431–1443.

Hampton, J. A., Estes, Z., & Simmons, S. (2007). Metamorphosis: Essence, appearance and behavior in the categorization of natural kinds. *Memory & Cognition, 35,* 1785–1800.

Hampton, J. A., & Gardiner, M. M. (1983). Measures of internal category structure: A correlational analysis of normative data. *British Journal of Psychology, 74,* 491–516.

Johnson, K. (2004). On the systematicity of language and thought. *The Journal of Philosophy, 101*(3), 111–139.

Jönsson, M. L. (2015). Overextension in verb conjunctions. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 41*(6), 1917–1922.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kim, S., & Murphy, G. L. (2011). Ideals and category typicality. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 37*(5), 1092–1112.

Kittur, A., Holyoak, K. J., & Hummel, J. E. (2006). Ideals aren't always typical: Dissociating goodness-of-exemplar from typicality judgments. In R. Sun & N. Miyake (Eds.), *Proceedings*

of the Twenty-eighth Conference of the Cognitive Science Society (pp. 429–434). Mahwah, NJ: Erlbaum.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: The University of Chicago Press.

Landau, B. (1982). Will the real grandmother please stand up? The psychological reality of dual meaning representations. *Journal of Psycholinguistic Research*, *11*(1), 47–62.

Larochelle, S., Richard, S., & Soulières, I. (2000). What some effects might not be: The time to verify membership in "well-defined" categories. *Quarterly Journal of Experimental Psychology, A, 53,* 929–961.

Malt, B. C., & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory & Cognition, 10,* 69–75.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition, 6*(4), 462–472.

Murphy, G. L., & Andrew, J. M. (1993). The conceptual basis of antonymy and synonymy in adjectives. *Journal of Memory and Language, 32,* 301–319.

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, 9*(1), 35–58.

Osherson, D. N., & Smith, E. E. (1982). Gradeness and conceptual combination. *Cognition, 12* (3), 299–318.

Osherson, D. N., & Smith, E. E. (1997). On typicality and vagueness. *Cognition, 64,* 189–206.

Pelletier, F. (2017). Compositionality and concepts—A perspective from formal semantics and philosophy of language. In J. A. Hampton & Y. Winter (Eds.), *Compositionality and concepts in linguistics and psychology* (pp. 31–94). Berlin: Springer.

Pullum, G. K., & Scholz, B. C. (2007). Systematicity and natural language syntax. *Croation Journal of Philosophy, VII*(21).

Rey, G. (1983). Concepts and stereotypes. *Cognition, 15,* 237–262.

Rice, S. A. (1992). Polysemy and lexical representation: The case of three English prepositions. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 89–94). Hillsdale, NJ: Erlbaum.

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

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192–232.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

Smith, E. E., & Sloman, S. A. (1994). Similarity-versus rule-based categorization. *Memory and Cognition, 22,* 377–386.

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

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

Storms, G., Ruts, W., & Vandenbroucke, A. (1998). Dominance, overextensions, and the conjunction effect in different syntactic phrasings of concept conjunctions. *European Journal of Cognitive Psychology, 10,* 337–372.

Werning, M., Hintzen, W., & Machery, E. (2012). *The Oxford handbook of compositionality*. Oxford University Press.

Wisniewski, E. J. (1997). When concepts combine. *Psychonomic Bulletin & Review, 4,* 167–183.

Wright, C. (1995). The epistemic conception of vagueness. *The Southern Journal of Philosophy, 33,* 133–160.