Principles for Concept Combination and Negation

Paper for the course Conceptual Semantics 2008, Matthijs Westera, 3021211

1. Introduction

The uniquely human capacity to read each others' minds resulted in linguistic communication, but also in a process called *iterative learning* (Tomasello 1999). This process gave rise to the *learning bottleneck* as an important constraint on language evolution (Jäger and Van Rooij 2007). Improved learnability in turn led to *compositional* languages becoming our dominant way of communicating with each other. In a compositional language, humans are able to combine expressions with existing meanings to refer to new meanings, and they can do so in a variety of ways, most of which are still not fully explained or understood. Many existing theories, developed to deal even solely with modifier-head combination (an adjective combined with a noun), encounter many problems and exceptions.

STRIPED APPLE, WHITE MAN, GIANT MIDGET, STONE LION, and PET FISH are only a few examples of concept combinations that are problematic, for reasons that will be made explicit throughout this paper. Partee and Kamp (1995), who tackled compositionality within prototype theory using supervaluations, blame all these problems on *context effects* and devised three principles, a kind of cognitive laws, to guide the dealing with such effects. How, exactly, these effects are dealt with according to those principles remains rather vague.

In this paper we will argue that the problems Partee and Kamp are trying to deal with are of very different sorts, and labeling them all as "context effects" does not resolve much. In section 2 we discuss their approach and evaluate two of their principles: the *head primacy principle* and the *non-vacuity principle*. We will then abandon their approach and distinguish between *contrast class effects*, treated in section 3, and effects due to *incompatibility* and *vacuity*, treated in section 4. In section 5 we will consider the effect of negation on concept combination and formulate for future research a testable hypothesis, the *postponed negation effect*, to establish or undermine the conceptual spaces framework for compositionality.

In the end we will have formulated five principles, which can be regarded as the cognitive building blocks of concept combination. Those are the *focus principle* and *structure reusing principle* to account for contrast class effects (section 3), the *expressivity principle* and the *domains economy principle* to deal with incompatibility and vacuity (section 4) and the *negation principle* to capture the essence of negation (section 5). Each of these principles is presented as a natural result of language evolution. Our arguments throughout this paper will apply specifically to the conceptual spaces approach[1], but the underlying principles should apply equally well to other theories of concept combination.

2. "Context effects"

In most of the problematic concept combinations mentioned in the introduction, problems arise because one concept modifies the other. Either the noun affects what the adjective means (e.g. "white man", "skillful surgeon", "midget giant") or the adjective modifies the meaning of the noun (e.g. "fake election", "former president", "stone lion"). But also simple intersective concepts (CARNIVOROUS, STRIPED[2]) pose a problem for concept combination,

[1] For a good overview we refer to Gärdenfors (2000).
[2] Concepts are called intersective if the extension of the combined concept is the intersection of both concepts' own extensions.

mainly in prototype theory, as noted by Osherson and Smith (1981). They observed that if we regard the composition of intersective concepts as fuzzy intersection, the typicality of a striped apple for the concept STRIPED APPLE can only ever be as high as either the instance's APPLE typicality or the instance's STRIPED typicality. This violates the *conjunction effect* found in human typicality judgments, that for example a striped apple is often more typically a STRIPED APPLE than an APPLE, as described by Osherson and Smith. Partee and Kamp (1995) cannot deal with this effect even after exchanging fuzzy logic with the more appropriate supervaluation theory. They simply call all these difficulties *context effects*.

Partee and Kamp explain most context effects as the result of a competition between two principles: the *head primacy principle*, claiming that the modifier is interpreted relative to the head context, and the *non vacuity principle*, claiming that at all times we should try to interpret concepts as actually meaning something. We will first illustrate the head primacy principle. Consider the concept WHITE MAN. The context of the man (we will make the word "context" more concrete soon) somehow alters the concept WHITE such as to denote the pinkish skin color that is actually quite unlike the white prototype. A more dramatic modifier-adjustment occurs in the concept PET FISH, where many of the typical PET-attributes (fluffy coat, has a name, sleeps on a mat) are dropped. Both examples are in accordance with the head primacy principle, here stated more formally:

> **Head primacy principle**
> In a modifier-head structure, the head is interpreted first, and then the modifier is interpreted relative to the local context created by the interpretation of the head. (Adapted from Partee and Kamp 1995).

With the concept STONE LION, something different is happening. If we apply the head primacy principle, we obtain a concept with an empty extension (there are no stone beasts that roar and eat meat). There is no way to adjust STONE such that the meaning of STONE LION is not empty. So we have a conflict between the head primacy principle and the non-vacuity principle, stated here more formally:

> **Non-vacuity principle**
> In any given context, try to interpret any predicate so that both its positive and negative extension are non-empty. (Adopted from Partee and Kamp 1995).

Enforced by the non-vacuity principle, the head primacy principle is abandoned and instead of the modifier, the head itself is adapted, in this case LION is changed into OBJECT IN THE SHAPE OF A LION.

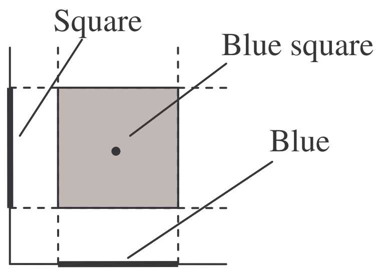
The non-vacuity principle makes sense from an evolutionary viewpoint. It is entirely in accordance with Grice's maxims to only communicate relevant and true information.[3] But the head primacy principle seems to us rather an ad hoc solution to account for context effects of a too large variety. In both "white man" and "pet fish" the meaning of the adjective is altered by the noun, but in "pet fish" this change is far bigger and of a different sort. Also the *kind* of adjustment in PET FISH seems to be different from that in STONE LION or WHITE MAN, but nothing of that sort is mentioned by Partee and Kamp. Furthermore, the conjunction effect described above is not dealt with by these two principles and is given an ad hoc solution by Partee and Kamp, which we will omit here.

[3] In this paper we will consider Grice's maxims as a solid evolutionary explanation. Of course, Grice's maxims in turn can be explained by e.g. social status (Jäger and Van Rooij 2007).

In this paper we will distinguish between *contrast class effects*, which explain the conjunction effect and modifier adjustment in combinations like WHITE MAN, and effects due to *incompatibility* and *vacuity*, which explain (and distinguish) PET FISH and STONE LION. Both principles mentioned so far will be abandoned, although the non-vacuity principle will appear again later in a generalized form.

3. Contrast class effects

In a conceptual space, properties are regions in conceptual domains. So to account for intersective modifier-head composition, we need to translate the union of all properties to some geometrical operation. To do so we need to distinguish two cases. In the simplest case the two concepts are fully independent. Arguably, BLUE and SQUARE are such concepts[4]. Warglien and Gärdenfors (2007) call this *direct composition*. In the second case the modifier and the head share one or more domains. As Warglien and Gärdenfors noted, what we are looking for in the case of direct composition is the *Cartesian product* operation, illustrated below.



We will assume that the similarity and typicality measures are simple monotone function of the distance in the space. If we use a simple, unweighted distance measure, we can see that all distances increase because the Cartesian product increases space itself (e.g. from 1D to 2D). This results in every instance of BLUE SQUARE being further away from the BLUE SQUARE prototype than it was from either the BLUE prototype or the SQUARE prototype in the original spaces.[5] In other words, many blue squares would be less typically a blue square than a square, which violates the conjunction effect.

Smith and his colleagues (1988) argued that a color concept like BLUE puts emphasis on the color domain of the concept it combines with. Important here is the notion of a *contrast class*. Within the fruit domain, the contrast class of apples are bananas, oranges, pears, with the important domains of comparison being shape, taste and color. As we go from APPLE to RED APPLE, the contrast class changes from a variety of fruits to only green and yellow apples. As a consequence, color becomes the most important criterion in the similarity measure. Smith and Osherson model this with a voting system, a color concept moving votes to the color domain, thus making it more important in comparison.

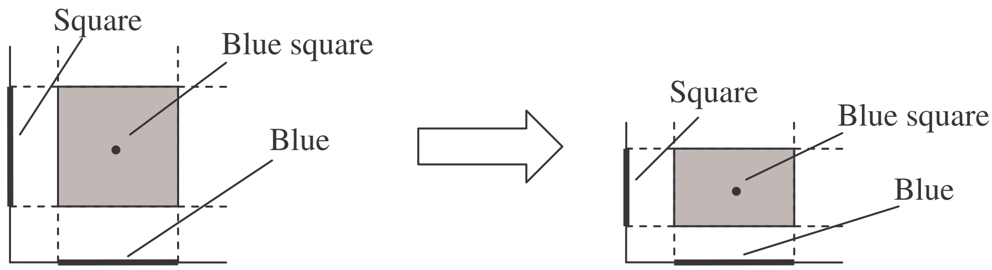
> **Focus principle**
> In modifier-head composition, the importance of the *contrast class* domains in the similarity measure is increased relative to the other domains.

---

[4] It can be argued that shape concepts like SQUARE already occupy a color domain, possibly with some default color white or black.

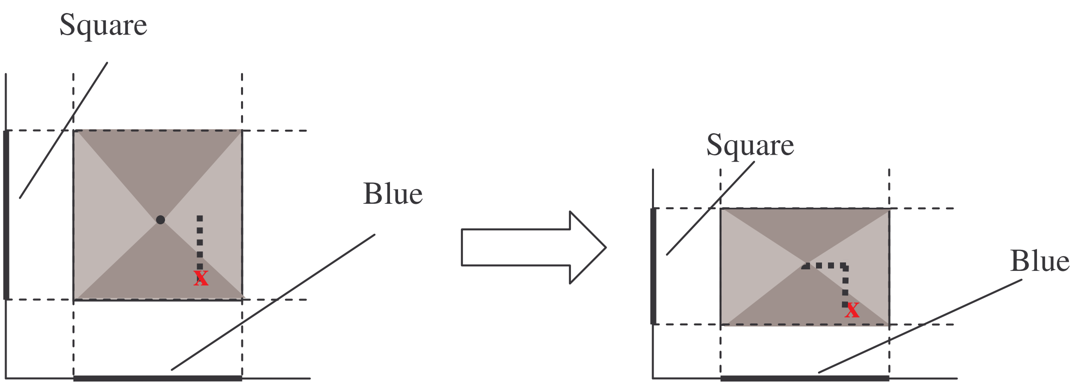
[5] If, instead of refering to the original spaces, we look at the distance to the BLUE or SQUARE prototype in the actual 2D space, things are just as bad, as we will soon see in a slightly different fashion.

In the conceptual spaces framework, the same effect can be achieved by assigning the color domain a higher weight in the distance function than the other domains of the concept. Since we wish to decrease the distance between points in the space, to incorporate the conjunction effect we should not increase the color weight, but decrease all other weights. In other words, the conceptual space should be *compressed* in all domains except the ones of the contrast class.



In the example, it is the weight of the shape domain (where the concept SQUARE resides) that is decreased. The conceptual space is thus compressed with respect to the shape domain, but not with respect to the color domain.[6] If this compression is too little, then the conjunction effect is only accounted for with respect to squares that are typically blue. If this compression is too strong, then the conjunction effect applies even to squares that are not even blue (such that e.g. a red square is more typically a BLUE SQUARE than a SQUARE). This is of course wrong.

To decide on the rate of compression, we have to find the color x, call it the *critical instance*, such that the concept X SQUARE is just as typically a BLUE SQUARE as it is a SQUARE. In other words, all instances that are less blue than the critical instance will be judged more typically a SQUARE than a BLUE SQUARE. We compress the shape domain in such a way that the distance of x to the BLUE SQUARE prototype in the new space is the same as the distance of x to the SQUARE prototype in the original space. This is depicted below for the Manhattan distance measure.[7]



The space much be rescaled such that the dotted line in the right space (the two segments together) is just as long as the dotted line in the left space. Obviously the critical instance must be more typically BLUE than SQUARE (i.e. in the shaded areas of the region), otherwise no rescaling of the shape domain can ever reduce the distances enough. In addition, the more typically BLUE we take our critical instance, the smaller the rescaling required to make all

[6] This means that for blue things that are prototypically square, the BLUE SQUARE typicality can only be as large as, or smaller than, the SQUARE typicality.

[7] The argument, including the equations yet to come, generalizes to any monotone distance measure, provided the equations can still be solved (i.e. linear in the weights).

less-blue instances more typically a SQUARE than a BLUE SQUARE. In extreme, if we take our critical instance to be a BLUE prototype, no rescaling is done and the conjunction effect would not be dealt with.

More formally, assuming the distance measure is a weighted sum of the distances in each domain, we need to solve the following equation for the weight w_i':

$$w_i \cdot d_i(x, p_i) = w_1 \cdot d_1(x, p) + \ldots + w_i' \cdot d_i(x, p) + \ldots + w_n \cdot d_n(x, p)$$

Where x is the critical instance and w_i' is the new weight of the i^{th} domain, the weight responsible for the compression. The weights $w_1 \ldots w_n$ belong to the uncalibrated distance measure, and $d_1(x,p) \ldots d_n(x,p)$ are the domains' distance functions. Solving it for the weight w_i' we wish to find gives:

$$w_i' = \frac{w_1 \cdot d_1(x, p) + \ldots + w_{i-1} \cdot d_{i-1}(x, p) + w_{i+1} \cdot d_{i+1}(x, p) + \ldots + w_n \cdot d_n(x, p) - w_i \cdot d_i(x, p_i)}{-d_i(x, p)}$$

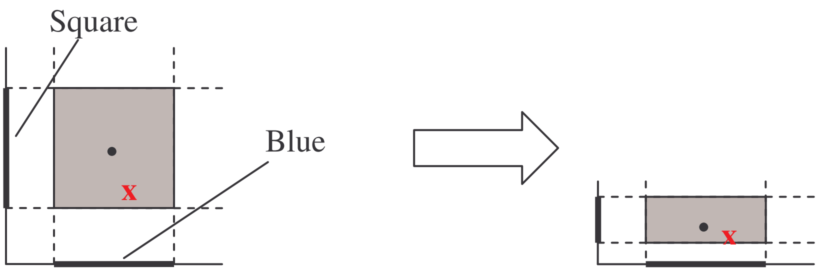
In words, choose your critical instance x, take the distance between x and the prototype disregarding the i^{th} domain, and divide it by the distance to the i-prototype minus the distance to the combined concept's prototype. To make it easier to grasp this procedure, let us consider the example of the blue square again. For simplicity we consider the original weights to equal 1. We need to solve the following equation, where c and s refer to the color and shape domains respectively:

$$d_s(x, p_{Square}) = d_c(x, p_{Blue}) + w_s' \cdot d_s(x, p_{Square})$$

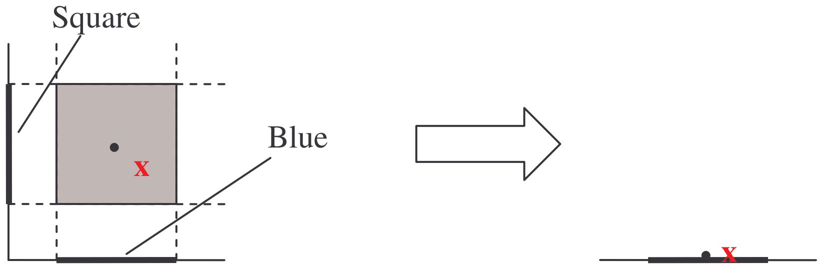
$$-w_s' \cdot d_s(x, p_{Square}) = d_c(x, p_{Blue}) - d_s(x, p_{Square})$$

$$w_s' = \frac{d_c(x, p_{Blue}) - d_s(x, p_{Square})}{-d_s(x, p_{Square})}$$

Remember that the critical instance is the instance for which we wish to have the same BLUE SQUARE typicality and SQUARE typicality (such that all instances that are less blue than the critical instance will be judged more typically a SQUARE than a BLUE SQUARE). A constraint was that $d_c(x,p_{Blue}) \leq d_s(x,p_{Square})$. We can see that if we take our critical instance x to be prototypically BLUE, w_s' will equal 1 and there will be no rescaling. Consider now as a critical instance the x such that $d_c(x,p_{Blue})=2$ and $d_s(x,p_{Square})=3$. The new shape weight w_s' is then given by (2-3)/-3=1/3. In other words, the shape domain is compressed to 1/3 of its original size. This results in the conjunction effect appearing for any instance that is more blue than the critical instance.

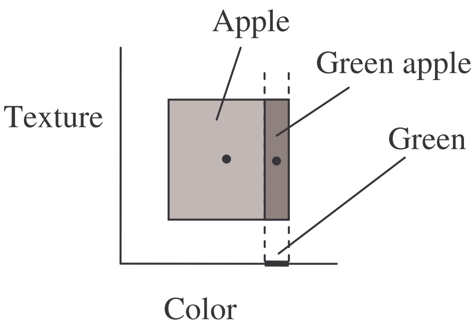


Now consider an instance *x* to be as typically blue as it is square (i.e. on a diagonal of the BLUE SQUARE concept). There is no way to rescale the shape domain such that *x*'s BLUE SQUARE and SQUARE typicality are the same but to remove the shape domain altogether. We indeed see that in that case solving the equation gives w_s'=0, leading to the shape domain being ignored entirely in the distance measure.



We have shown that rescaling based on a critical instance gives us the results one would expect and allows us to deal with a conjunction effect of arbitrary strength. We will now look at concept combination where the modifier and the head have domains in common. In the general case, the regions of a modifier replace the regions of the head concept in all shared domains.

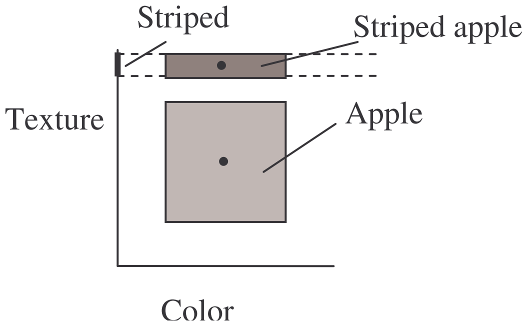
Consider the concept GREEN APPLE. Obviously, color is already a domain for the concept APPLE, so an increase in the dimensionality of the space, via the Cartesian product, is not required. Instead, the region of the concept is simply narrowed down to a more specific sub-region in the same space. As with the BLUE SQUARE example, here too the conceptual space should be compressed to put an emphasis on the color domain. We will omit it for now, but we will need it again soon.



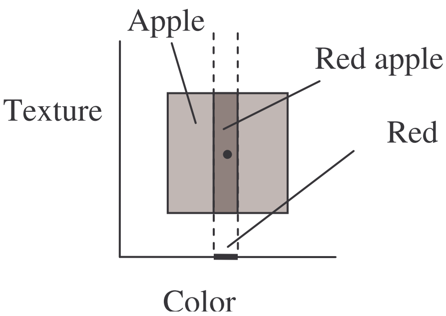
Because our space itself does not increase, the conjunction effect is successfully accounted for already. We can see in the illustration that all instances in the dark grey area (i.e. all green apples) will be closer to the GREEN APPLE prototype than to the APPLE prototype. So any green apple will be more typically a green apple than it will be an apple – which is as it should be.

Osherson and Smith (1981) discovered that the conjunction effect is larger for incompatible concepts[8]. This follows from the conceptual spaces approach naturally. For example, the region for STRIPED APPLE is far away from the regular APPLE region (see the illustration below). Because of that, all striped apples will be much closer to the STRIPED APPLE prototype than to the APPLE prototype – more so than was the case with GREEN APPLE and APPLE. Obviously, the more incompatible the concepts, the larger the conjunction effect.

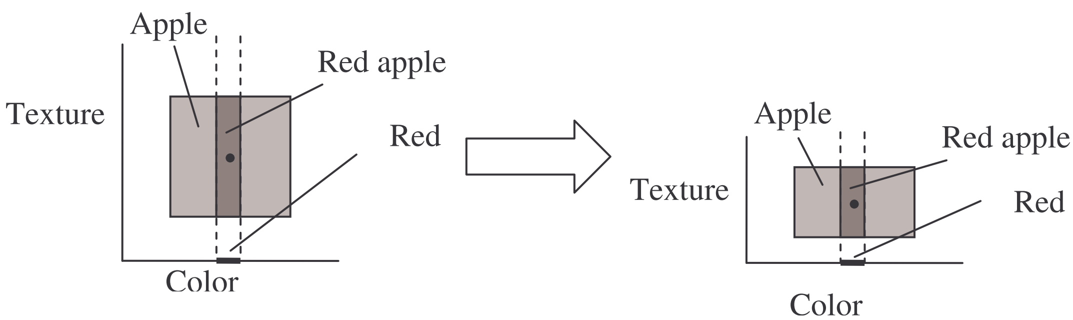
[8] This notion of incompatibility is different from the one treated in section 4. I will nevertheless use it in this section in the sense intended by Osherson and Smith.



At the other extreme, instead of incompatible, the combined concept's prototype could be very close to the original prototype, e.g. RED APPLE to APPLE. Assuming, for the sake of argument, that red apples are among the most typical apples, we find that any apple is just as typical a RED APPLE as it is an APPLE.[9] It seems that in this extreme case we cannot account for the conjunction effect. (It is in accordance with Osherson and Smith that the conjunction effect should be weaker in this case, but it should be occurring nevertheless).



To deal with the conjunction effect here, we rely on compression as before. In the case of the BLUE SQUARE concept, we had to adjust the weighted distance measure to decrease the distances. Here, too, the combination of RED with APPLE emphasizes the color domain, leading to a compression of the other domains (texture, in this illustrative case). The amount of compression depends on the critical RED APPLE instance we choose.

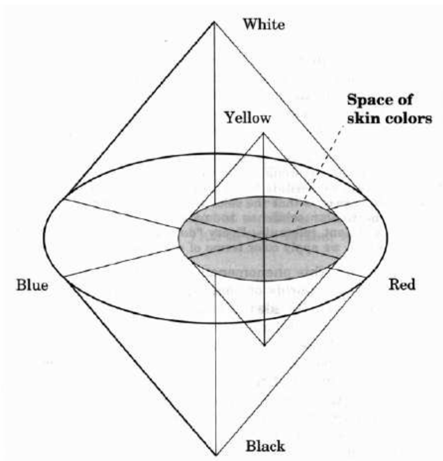


We have seen that rescaling is, in principle, not required to account for the conjunction effect in GREEN APPLE and STRIPED APPLE, but we will generalize and assume that rescaling always occurs. In the GREEN APPLE and STRIPED APPLE case, even though there is a conjunction effect already without compression, the ability to pick a critical instance and still compress the space allows us to decrease or increase the conjunction effect to our likings. We have thus shown that conceptual space compression is all we need in order to account for the conjunction effect in conceptual spaces for both independent and dependent concepts. Furthermore, because of the conceptual spaces architecture the conjunction effect is automatically stronger the more incompatible the combined concepts are.

9 If this is not convincing, consider the concept TYPICAL APPLE, and see that any apple is just as typical a TYPICAL APPLE as it is an APPLE.

This compression leads us to the question of how to obtain this critical instance the rescaling is based on. In the RED APPLE case, the critical instance can be seen as the apple marking the borderline between when to categorize something as a RED APPLE and when to speak simply of an APPLE. Any apple that is less red than the critical instance will be categorized as an APPLE instead of a RED APPLE. Where this critical instance resides, i.e. far away from or close to the RED prototype, depends on the classical *cognitive economy* compromise between information value and computational effort (e.g. Jäger and Van Rooij 2007). The utterance “red apple” carries more information than the utterance “apple”, but it is also longer and more complex. At some point, a particular apple is no longer red enough to make this information gain on its color worth the effort, and the utterance “apple” is preferred. Of course the information gain is fully dependent on the context of uttering. If you want someone to take a particular apple, it will be worthwhile to utter the longer sentence “take the red apple” even if it is only slightly reddish, as long as the apples around it are all green.

The latter example touches on a second effect of contrast classes on the structure of the conceptual spaces, which has been emphasized by Gärdenfors (2000). Not only becomes the color domain more important as a result of the contrast class change, the interpretation of the color concepts themselves also change depending on the contrast class. Obviously, the concepts mentioned before, like LONG and GIANT, are for their interpretation fully dependent on the contrast class of the concept they combine with. But also concepts that do have their own prototype may change radically when combined with other concepts. The contrast class of human skin colors (ranging, say, from pinkish to brownish) changes the meaning of the color words “white”, “yellow”, “red” and “black” to denote skin colors, which are in fact far from the prototypical colors white, yellow, red and black. What happens is that the structure of the whole color space as we know it is reused for the subspace of human skin colors. The picture below (taken from Gärdenfors 2000) illustrates this.

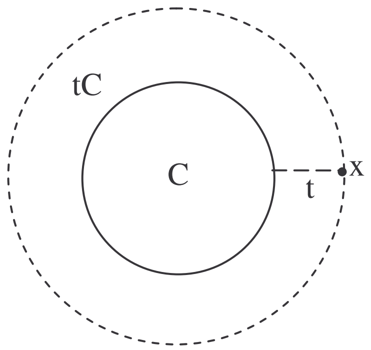


Before we formalize this notion of structure reusing, we will formulate this in another principle and see where it stems from.

> **Structure reusing principle**
> In modifier-head composition, the structure of the conceptual space of the modifier is reused as much as the head’s contrast class allows.

This principle is a natural result of cultural constraints on the evolution of language. As Jäger and Van Rooij (2007) mention, even in a new context with only minor similarities to the usual context one often has a larger chance of being understood by reusing existing words than by inventing a new word for the occasion. The structure reusing principle captures this idea, and presumes an easy tool to reuse existing concepts in novel contexts. Also, reusing structure increases, as does compositionality itself, the learnability of a language.

This easy tool exists, as Warglien and Gärdenfors (2007) proposed. We will just briefly treat it here, without going into the technicalities. By using the mathematical notion of *gauge*. Informally, the gauge of a point x is the smallest expansion t of a concept C such that x is in the expanded concept tC. The gauge notion is illustrated below.



If you now take any point x, you can take its gauge for one concept C and divide it by its gauge for another concept D. If you then multiply x by this value, you obtain what is called the *radial projection x'* of x from C into D. This works for any two concepts that have a common point (the *origo)* and that are *convex* and *compact*. These last two prerequisites will come back in section 5, because negations are often non-convex regions.

4. Expressivity effects

We will now discuss what are often called incompatible concept combinations, such as PET FISH and STONE LION. We think that these two combinations are of a rather different sort. In the first case, PET FISH, the modifier occupies many domains that are not included in the head's conceptual space (e.g. coat fluffiness, the sound it makes), i.e. the domains are truly *incompatible*. In the second case, there is no incompatibility between the domains, but rather there is *vacuity*: the combination results in a concept with an empty extension (there are no stone beasts that roar and eat meat, after all). The criteria of compatibility and non-vacuity can be described in the following principle, which is a generalization of Partee and Kamp's non-vacuity principle (1995).

> **Expressivity principle**
> In any given context, try to interpret any predicate so that it is expressive, i.e. a) an extension can be assigned to it and b) both its positive and negative extension are non-empty.

We can see that the expressivity principle flows naturally from Grice's maxims to always provide meaningful information. According to these maxims, one should always provide relevant information – so no concepts with an empty negative extension (tautologies like FOUR-LEGGED LION). In addition one should never talk about things that do not exist, i.e. concepts with an empty positive extension (like STONE LION). The principle in fact reflects on these maxims, saying that you should always assume that the other behaves according to the maxims. So the non-vacuity principle does not prohibit that we use the words "stone lion". Instead, in case one does say "stone lion", the principle forces us to try to interpret it as something expressive, for instance as a STONE OBJECT IN THE SHAPE OF A LION.

To achieve non-vacuity and compatibility, the conceptual spaces could be restructured easily by just adding or removing arbitrary domains. We argue that this restructuring happens economically, i.e. according to the following principle:

> **Domains economy principle**
> To resolve incompatibility or vacuity, domains can only be dropped from a conceptual space, never added (unless, of course, if there are no shared domains[10]) – and as little as possible.

The reasons for this principle are straightforward. If there is no constraint on removing and adding domains in order to resolve incompatibility, the end is simply lost. It would not be economical and communication would likely fail. We cannot think of any examples against this principle, i.e. concepts that, when combining, *gain* extra domains. Also, exceptions to this rule are not automatically counterexamples to the domains economy principle. It is very well possible that, once concept incompatibility is resolved, new domains are associated with the combined concept.[11] A good counterexample should exclude the latter associative step from happening, as this principle is solely about the combination step.

Warglien and Gärdenfors (2007) seem to argue that the decision on which conceptual space to restructure (and how) involves world knowledge. With respect to the STONE LION example, they claim that "living is presumed by many of the domains of LION. These domains, like sound, habitat, behavior, etc, can thus not be assigned any region at all". In other words, domains like habitat are dropped because the modifier STONE cannot be united with LIVING, and living is a precondition for that kind of domain. They seem to presume that domains themselves can have properties or preconditions, and can be related to one another, but fail to make explicit how this works. Warglien and Gärdenfors remain too vague to be able to really grasp what they mean, but attaching terms like 'presume' to domains suggest that what is happening is something complex, involving a lot of world knowledge, causality, etcetera. But in fact this is not the case. We think that the mechanism is a natural consequence of the domains economy principle and the expressivity principle, as we will here show for both concepts.

PET FISH. Assume, for the sake of argument, that PET occupies several domains unrelated to FISH. Related to PET are concepts like FLUFFY, CAGE, the SOUND it makes (barking, meowing) and that it has an OWNER. A modifier should replaces regions in the domains of the head, but this cannot be done for each of the modifier's domains, i.e. the concepts are truly incompatible. The only way we can make the conceptual spaces compatible is by dropping some domains of the PET concept, leaving in only the domains for CAGE (possibly restricted, by means of the contrast class, to BOWL) and OWNER.

STONE LION. When someone talks about a stone lion, initially the concept's extension is empty. In order to make the concept meaningful, some domains have to be dropped. In the combination STONE LION, STONE might only occupy a single domain (e.g. material), so restructuring the modifier has no future. Instead we start dropping domains from the LION

---

[10] If we recall the simplest example so far, BLUE SQUARE, we now see that this combination in fact illustrated completely mismatching domains. The concepts BLUE and SQUARE did not share any regions at all. In order to combine the two concepts, the head SQUARE had to be expanded to include the color domain – it was a last resort.

[11] In fact, this kind of associating may be another important building block for human cognition, but one that is irrelevant to a theory of concept combination.

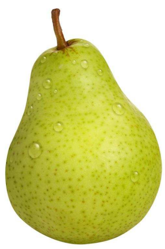
conceptual space. The only way to make the positive extension of STONE LION non-empty is to remove domains like sound (there are no stones that roar), habitat (there are no stones with a habitat), and behavior (there are no stones with a behavior). There are stones with a particular shape, so the shape domain remains intact.

It is worth mentioning here, following Warglien and Gärdenfors (2007), that radial projection allows one to map any two conceptual spaces onto one another (as long as they are convex and compact). This allows *metaphorical composition*, where the structure of one domain is transferred onto that of another domain, taking all the concepts with it. Metaphorical composition seems to be driven by the expressivity principle again: if the concepts are incompatible or if the combined concept is empty, try whether it is a metaphor! In Western culture, often the money domain is projected onto the time domain, allowing us to use many time-as-money metaphors, such as “spend time”, “afford the time” and “it costs time”. Warglien and Gärdenfors show how the vertical/horizontal space can be transferred to a status/time space, allowing one to speak of the peek of one’s career. Because it is outside the scope of this paper we will not go into this any further. It is clear anyhow that metaphorical transfer adds yet another powerful tool to concept combination.

5. Negations

In fuzzy logic and supervaluation approaches towards concept combination, negation is often seen as a simple typicality inverter, e.g. $C_{notA} = 1 - C_A$. However, there are very common exceptions to this rule. For example, a chair has a very low APPLE typicality, but that does not mean it is a very typical non-apple. There are two problems that need to be dealt with in order to account for negation appropriately. First, what domains do negated concepts reside in? Second, how can a region that is non-convex, which is often the case with negated concepts, have a prototype and a typicality measure? We will first discuss and attempt to answer the first question.

The concept NOT combines with other concepts to exclude them from the meaning conveyed.[12] The remaining concept, e.g. NOT AN APPLE, is obviously a very broad one. But how broad is it? Does its reference include pears and strawberries, tables and chairs, animals? All those things are most definitely not apples. Strangely enough, a chair is much more unlike an apple than a pear is, yet the label “not an apple” strikes us as much weirder when attached to a chair than when attached to a pear. It is certainly not impossible, for a chair is definitely not an apple, but it is not a normal use of NOT.



Not an apple.

?Not an apple.

It is fertile to look at the function of negation in day to day situations. According to Grice’s maxims, we are obliged to (1) always give sufficiently meaningful information and (2) not

[12] For simplicity we will call NOT a concept, although the word “operator” might be more appropriate.

give any irrelevant information. Both of these maxims shed light on when to or not to use the word "not".

Ad (1): The concept NOT is only sufficiently meaningful when it is used to exclude from consideration one of a handful of valid alternatives. It is not meaningful at all when it leaves open thousands of alternatives. For example, in a fruit-guessing game, "it's not an apple" is a meaningful hint. But in a food-guessing game, "it's not an apple" is not very useful.
Ad (2): The concept NOT is only relevant when it is not immediately clear that a particular instance or class is not to be considered a valid candidate. The label "not an apple" is completely irrelevant when attached to a chair, for even for infants it is obvious that a chair is not an apple. But when attached to a round pear, it might be very relevant (at least for infants).

Summing up, the word "not" is only used appropriately, i.e. in accordance with Grice's maxims, when applied to cross out one valid candidate out of a limited number of candidates. This is only the case when all candidates reside in the same conceptual domains, e.g. all fruits, all animals or all tools. For that reason we think it safe to assume that NOT combines with another concept by taking the universal concept in the concept's domains, minus the concept itself.

In this sense, the concept NOT AN APPLE does not include such things as chairs and animals. It is only when explicitly considering a sentence like "this chair is not an apple" (or seeing a chair labeled as such) that we are forced to expand the conceptual space under consideration to a more general one so as to include both concepts APPLE and CHAIR, and see indeed that a chair is not an apple. This required change in conceptual view explains the strangeness of a chair labeled "not an apple". [13]

Such a shift in conceptual spaces is not unique to cases where something is very obviously not an apple. Consider, this time, a fake apple labeled as "not an apple". Only when expanding the conceptual view to include fake apples (e.g. with an added domain for decorative value) can the sentence "this is not an apple" seen to be true. The more substantial the required conceptual space change, the bigger is the alienation. A good example is the famous artwork by Magritte with the label "Ceci n'est pas une pipe.". Seeing the truth of the sentence requires a large change of perspective, expanding the object-relevant conceptual domains with those relevant for pictures of objects.

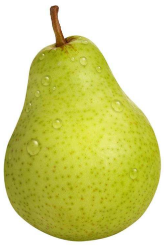
We can summarize these findings as following principle, the evolutionary origins of which we have already explained:

[13] In fact these feelings of strangeness are not unique to out-of-domain constructions with negation. They seem to occur whenever a Gricean maxim, or any communicative habit, is violated. "Would you like tea or coffee?" "Yes." Many jokes seem to rely on such violations.

Negation principle

The negation of a concept C resides in the same conceptual domains as C itself, i.e. it is the complement of C in the same conceptual space, *unless a different, more general conceptual space is enforced by the context*.

The negation principle tells us what region corresponds to the negation of a given concept. A consequence of this is the second problem we briefly described above, that negated concepts are often non-convex regions. The lack of convexity entails the lack of a region center, and thus the lack of a straightforward prototype altogether. Certainly it is questionable whether such concepts have prototypical members. For example, try to consider the prototypical NON-APPLE. Is it a pear? A strawberry? A chair?



?Typically not an apple.

?Typically not an apple.

??Typically not an apple.

Smith and his colleagues (1988) have described typicality effects for negated concepts, revealing that negation was roughly order-inverting. It remains a question for further research whether such typicality effects occur in every concept (and whether typicality is determined by a prototype). It might be the case that this only happens when staying within a narrow domain (e.g. the fruit domain for NON-APPLE).

We will now discuss the combination of concepts when negation is added to the picture. As we described before, many concept combinations involve contrast classes. In the concept WHITE MAN, the color contrast class of MAN modifies the concept WHITE. This requires a projection of the MAN concept onto the color domain. Through the projection, we distill the fact that all men have a color ranging between pink and brown. The color spindle and the concept WHITE within it are then rescaled to match the skin colors, via a radial projection. There are two points at which this might go wrong when negation is added to the mental lexicon. First, negated concepts lack useful contrast classes, which is a problem for concepts like WHITE NON-MAN. Second, negated concepts tend to lack convexity, which is a prerequisite for radial projection and therefore a problem for concepts like NON-WHITE MAN.

Let us start with the first problem. According to the negation principle, negated concepts stay within the same conceptual space domains, which in this case are the domains relevant to the HUMAN BEING concept. The concept NON-MAN includes any concept in this space that has one or more properties that are unlike men. Such properties include low voice, facial hair, the reproductive system, but also trivial things like having two legs and having a human-like skin color. Strictly speaking, a purple creature that has all properties except color in common with a man, is not a man (i.e. a non-man).[14] Because any color is acceptable, the color contrast

[14] This may seem odd, because the purpleness may be due to a disease. But we think that the reason for considering purple man-like creatures as men is due to world knowledge, not due to the mechanism responsible

class for NON-MAN, i.e. the projection of NON-MAN onto the color domain, is the full color spindle. As a result, WHITE will refer to prototypical white, instead of the desired pinkish skin color. So the color contrast class of NON-MAN is useless for adjusting WHITE. Instead, in modifier-head composition with a negated head we feel that the negated head's contrast class should be identical to the non-negated head's contrast class. In other words, to construct WHITE NON-MAN, the color contrast class of MAN is taken to adjust WHITE, and after that the two concepts are combined in the usual way. In this way, "white non-man" would mean roughly the same as "a creature with a human-like pinkish skin color that is not a man".

Consider now the second case, where the modifier is a negated concept, for example in NON-WHITE MAN. To rescale a concept like NON-WHITE to the human skin color contrast class, the color spindle and the NON-WHITE region within are *radially projected* onto the skin colors, a prerequisite of which is the convexity of the concept. Obviously, the concept NON-WHITE is not convex, so this radial projection is likely to fail. To solve this, like before, all the rescaling needs to be done before the negation is added. This means that first the negation-less concept WHITE is projected radially onto the skin colors, and then negation is applied to this modified WHITE concept. The resulting concept NON-WHITE MAN has exactly the intended extension: all men that do not have a light pinkish skin color.

In modifier-head composition where the two concepts involved only share part of their domains, such as PET FISH and STONE LION, similar problems occur when negation is brought into the picture. As described above, such concepts are composed by changing the conceptual space of either the modifier or the head, i.e. by structural reconfigurations of the domains involved. Consider the concept STONE NON-LION. Similarly to the projection of NON-MAN onto the color domain, the projection of NON-LION onto the shape domain results in a universal concept there - after all, non-lions exist in every shape. As a result, the concept STONE NON-LION would wrongly include stone objects in the shape of a lion – precisely the objects we would call "stone lions". To solve this, we employ a similar strategy as with the contrast classes. We enforce that the conceptual space restructuring is done based on the non-negated concept, and apply the negation only at the end of the procedure, resulting in the proper concept STONE OBJECT NOT IN THE SHAPE OF A LION.

Generalizing over all the cases treated here, we can make a prediction on concept combinations with negation.

> **Postponed negation effect**
> In modifier-head composition, negation is applied only after the necessary contrast class adjustments or conceptual space restructurings.

The postponed negation effect would make sense in light of the evolutionary origins of convex categories. The negations of convex concepts are themselves, in general, not convex. Not only do non-convex concepts create problems for concept combination using radial projection, non-convexity also no longer ensures a *meeting of minds* as described by Warglien and Gärdenfors (2007). Therefore it makes sense to postpone the loss of convexity due to negation as much as possible.

The postponed negation effect is a prediction. It flows naturally from the negation principle, for which we gave an evolutionary argument, and from the evolutionary advantages of

---

for negation composition. If we knew a whole *race* of such purple creatures to exist somewhere on a distant planet, surely they would be considered really unlike men.

convexity. However, we do not know whether this effect has in fact been observed in humans. We think that brain activity experiments could be done to investigate this. First, it must be investigated what brain activity corresponds to negation (not to reading the word "not", but to actually processing it and applying it to a concept). Then the brain must be scanned while participants are processing sentences or combining concepts. This admittedly optimistic sketch of an experiment would reveal whether the postponed negation effect is true, i.e. whether indeed NOT is always applied at the end. If this turns out to be not the case, we should seriously reconsider compositionality in conceptual spaces, and see what other ways – if any – the mind has found to deal with contrast classes and space restructuring.

6. Conclusions

We have first described some of what Partee and Kamp (1995) call 'context effects' and explained the head-primacy principle and the non-vacuity principle. We argued that their account of concept combination completely lacked precision and missed some important distinctions. A key distinction was made between contrast class effects, which include the conjunction effect and concepts like WHITE MAN, and effects due to incompatibility and vacuity, as in PET FISH and STONE LION.

Based on this distinction we formulated five principles that govern concept combination. These principles are the focus principle, the structure reusing principle, the expressivity principle, the domains economy principle, and the negation principle. The first two are responsible for contrast class effects and have been explained evolutionarily in terms of cognitive economy and commensurability/learnability constraints. The second two govern the restructuring of conceptual space due to incompatibility or vacuity and have been explained in terms of Grice's maxims and cognitive economy. The negation principle, finally, was explained again in terms of Grice's maxims and captures the way in which negation applies to concepts.

We have thus presented five cognitively and evolutionarily plausible principles that together account for a lot of the 'problematical' concepts and the conjunction effect. These principles shed light on concept combination, but can in principle be 'explained away' in terms of more elementary evolutionary principles. In addition, from the negation principle and the structure reusing principle, together with the general evolutionary advantage of convexity, we inferred the postponed negation effect, which provides a testable hypothesis for future research.

References

Gärdenfors (2000). "Concept combination: a geometrical model", pp. 129-146 in L. Cavedon, P. Blackburn, N. Braisby and A. Shimojima (eds) Logic language and Computation Vol 3, CSLI, Stanford, CA.

Jäger, G. and R. van Rooij (2007). Language structure: psychological and social constraints, *Synthese* **159(1)**, 99-130.

Osherson, D. and E. Smith (1981). On the adequacy of prototype theory as a theory of concepts, *Cognition* **9, 1**, 35-58.

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

Smith, E., D. Osherson, L. Rips and M. Keane (1988). Combining prototypes: a selective modification model, *Cognitive science* **12**.

Tomasello, M. (1999). The cultural origins of human cognition, Harvard University Press.

Warglien, M. and P. Gärdenfors (2007). Semantics, conceptual spaces and the meeting of minds, *manuscript*.