How to solve the conjunction fallacy?

A discussion of alternative approaches

MSc Thesis (*Afstudeerscriptie*)

written by

D.S. Harten
(born January 10th, 1984 in Amsterdam, the Netherlands)

under the supervision of **Reinhard Blutner**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
June 25th, 2008	Reinhard Blutner
	Frank Veltman
	Peter van Emde Boas

I'm dedicating this thesis to my grandmother 'oma Nel', who would have loved to read it.

Contents

List of Figures

Chapter 1

Introduction

Since many years, people have tried to understand and to model their own behavior and language. This has proven to be a difficult, yet rewarding task. Since the development of artificial intelligence and robots, there is an actual use for this modeling, besides the sheer happiness of understanding a part of our own existence.

One of the problems that we encounter when dealing with human language, is our way of handling concepts and creating new, complex concepts out of simpler ones. A concept can be seen as a mental representation [the Stanford Encyclopedia of Philosophy, 2007]. For example, let us consider a real life object 'chair'. This is something different from what we have in mind when we say the word 'chair'. Because when we are talking about it, we are using a mental representation and not the object itself. In other words, our mental representation will probably be a prototypical chair, where the real life chair does not need to be.

When we are creating complex concepts out of simpler ones, we are actually trying to describe the complex object we see, by combining two or more simple concepts that are already familiar to us.
This thesis will investigate the different possibilities to solve one of the problems that occur when we are trying to create complex concepts: the conjunction fallacy. The conjunction fallacy occurs when the combination of two concepts has a higher probability than the original concepts. This thesis will explore what research has been done through the years in this field. It will define different ways in which the fallacy can be interpreted and it will try to find a solution for the conjunction fallacy.

I have divided my thesis into three parts. The first part handles the different approaches to a solution for the conjunction fallacy using a 'classical' Boolean algebra.

The second part handles the more recent approaches that use a non-Boolean algebra and geometrical models.
Finally, the third part contains the conclusion and future work.

Part I

Concept combination using a Boolean algebra

Chapter 2

Prototype Theory and Fuzzy Set Theory

The Stanford Encyclopedia of Philosophy states that according to prototype theory, "a lexical concept C does not have definitional structure but has probabilistic structure in that something falls under C just in case it satisfies a sufficient number of properties encoded by C's constituents"[the Stanford Encyclopedia of Philosophy, 2007]. This means that every concept has at least one prototype that is the model of all the objects in a certain concept class.
For example: A sparrow is a very good prototype of the concept class 'bird'. It flies, has wings and feathers and lays eggs. Most birds are similar to the sparrow. However, it is possible that there are objects in a concept class, that do belong to that class, but are not very similar to the prototype. For example: a penguin cannot fly and does not resemble a sparrow, but it still is a bird.

Prototype theory is a widely used theory that is rather intuitive and is psychologically based [Sternberg, 2003]. Some people say it is founded by Plato [Plato]. He constructed his 'idea theory'. This theory states that everything in the real world is a copy of the original that is in the idea world. Just as every cookie is a copy of the cookie form. They do resemble each other, but are not completely the same.

Sternberg describes in his book 'Cognitive Psychology' [Sternberg, 2003] the two main approaches of prototype theory in psychology: prototype based cognition and feature based cognition[1]. Prototype based cognition means that an object is recognized as a certain concept by comparing it to a prototype. In feature based cognition, a concept is described by a 'simplest'

[1]For another clear explanation of this, see the article 'What some concepts might not be' by Armstrong, Gleitman and Gleitman [Armstrong et al., 1983]

set of properties that are individually necessary and sufficient to uniquely describe this particular concept.
The papers discussed in this article can be divided into these two groups as well. Osherson and Smith's theory has prototype cognition as a basis. Kamp and Partee use both types in their supervaluation theory. In the end I will discuss (among others) Gärdenfors and Franco, who have a more feature based approach.

2.1 Osherson and Smith's Prototype Theory

Osherson and Smith's interpretation of prototype theory states that a concept consists of a set of objects and that every concept has a best example, a prototype. Contrariwise to the definition of the Stanford Encyclopedia [the Stanford Encyclopedia of Philosophy, 2007], concepts in this theory can only have one prototype. Often this prototype holds on every property (every dimension) the average value of the same properties of all the other objects in the class.

There are, however, always objects that do not belong to only one class, but to two or maybe even more. This is exemplified in figure 2.1:

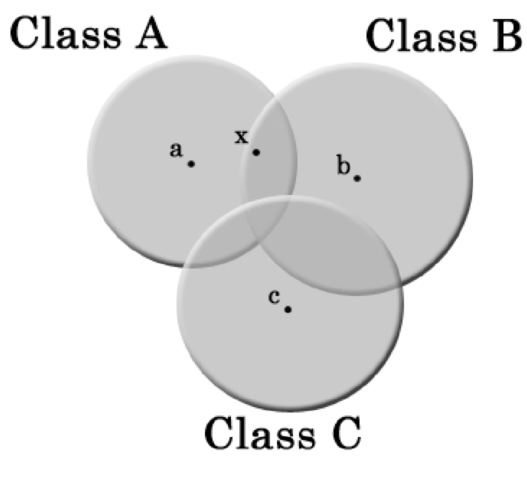


Figure 2.1: Fuzzy sets

Object a is the prototype of class A, object b is the prototype of class B, object c is the prototype of class C and object x can be both in A and in

B. This is an example where class A and B overlap. The boundaries aren't sharp, but rather vague[2]. These are often called *fuzzy set boundaries.*
In order to determine to which class object x will most likely belong, we have to determine the distance between x and the prototypes a and b. The prototype that is closest to x (in this case, a), is the prototype of the class that x most likely belongs to.

In order to do this, every set is represented as a quadruple $< A, d, p, c >$, where A is the set of already visible objects, d is the distance metric, p is the prototype of set A and c is the characteristic function.
The distance metric d combined with the set of objects A make a metric space, where [Osherson and Smith, 1981]

Definition 2.1. *$(\forall x \in A)$ $(\forall y \in A)$ $(\forall z \in A)$*

1. $d(x, y) = 0$ *iff* $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, y) + d(y, z) \geq d(x, z)$

The characteristic function c is the function that assigns a number from [0,1] to a certain object, to express the closeness to the prototype p. The closer an object is to a prototype, the higher is its value assigned by the characteristic function. This is expressed in the following expression:

Definition 2.2. *$(\forall x \in A)$ $(\forall y \in A)$*
$d(x, p) < d(y, p) \rightarrow c(x) > c(y)$

This results in that all objects are graded in their concept membership. As long as an object isn't a prototype, this function will assign a value between 1 and 0 to an object which denotes its closeness to the prototype and therefore denotes the grade of its membership to that concept.

This grading of membership causes rather vague boundaries of a concept. But in order to perform a concept combination, it is important to know to

[2]A question that arises when we use the word *vague*, is *What is vagueness?*. Kit Fine gives a rather nice definition in his article "Vagueness, truth and logic"[Fine, 1975]:

> "I take it to be a semantic notion. Very roughly, vagueness is deficiency of meaning. As such, it is to be distinguished from generality, undecidability, and ambiguity. These latter are, if you like, lack of content, possible knowledge, and univocal meaning."

This means that when something has a fuzzy boundary, this concept is not enough explained, not enough specified.

which class an object does (most likely) belong. In order to determine the membership, fuzzy set theory is used. In the next section we shall take a closer look at that.

2.2 Fuzzy Set Theory and the composition of prototypes

Classical set theory is based on one of the three fundamental laws of thought of Aristotle: the law of the excluded middle. This law states that for every proposition p, either p or $\neg p$ is true. Translated to classical set theory, this law holds that for every proposition p, p either is, or is not a member of a set S.

Around 1910, Jan Łucasiewicz extended this two-valued logic into a three-valued one, by adding an extra truth value: *possible*. So now, every proposition p is either a member or not a member of a certain set, or a possible member.

Lotfi A. Zadeh proposed in 1965 [Zadeh, 1965] fuzzy set theory, an extension of classical set theory.
He introduced a function that would give for each proposition the chance (between 1 and 0) that it was a member of a certain set. Fuzzy set theory was one example of an implementation of fuzzy logic. Zadeh also showed that fuzzy logic is just a generalization of classical logic.

Fuzzy set theory basically works the same as classical set theory, but here the characteristic function assigns not just a 0 or a 1 to an object, but a rational number between 0 and 1: $c(x) \to [0,1]$.

This ensures graded membership. The closer to 1 the value that the characteristic function assigns to a certain object is, the closer that object is to the prototype of that class and the greater the chance that the object does actually belong to that class. Only the prototype itself receives the value of 1.

It is also possible for an object to have a membership value for multiple classes. The object x from figure 2.1, will have a value for class A and for class B. But because x is closer to the prototype of A than to the prototype of B, the object will have a higher membership value for class A.

This is a very intuitive approach to implement prototype theory. In our

daily live we encounter many objects that do not obviously belong to a single class, but we 'put' them in the class that seems most likely to us at that moment.
A good natural example for this is a tomato. Many people consider this a vegetable and use (and eat) it as a vegetable. But officially it is a fruit.
Another example is a whale. This animal lives in the sea like a fish and looks like a fish, but is actually a mammal.

There are few 'natural examples' of concepts with fuzzy boundaries. There are, however, many more 'artificial examples': examples of man made objects that do not belong to just one class, such as a chaise longue (is it a bed or a chair?), a beanbag (is it a pillow or a chair?) and a soup plate (is it actually a plate, or is it a bowl?).

Let us now take a look at the result of using fuzzy set theory. I will explain the results with an example.

Example 2.1
A beanbag is a large bag, typically filled with polystyrene beads, used as a seat. Therefore, the value for $c_{Chair}(beanbag) = 0.8$. Because it is often used to sit on, but it is not a prototypical chair. The value for $c_{Cushion}(beanbag) = 0.5$, because it resembles a cushion, but is not used that way. The sentence 'A beanbag is a chair and a cushion', should receive a rather low value, but not zero. It is quite unlikely that something is both a chair and a cushion, but in the beanbag case, it is possible.

This is what we get by using fuzzy set theory[3]:

$$\begin{aligned} c_{Chair \cap Cushion}(Beanbag) &= min(c_{Chair}(Beanbag); c_{Cushion}(Beanbag)) \\ &= min(0.8; 0.5) \\ &= 0.5 \end{aligned} \tag{2.1}$$

The sentence 'A beanbag is a chair or a cushion' should receive a rather high value, because a beanbag is used as a chair and resembles a cushion. Therefore the chance that a beanbag is one of the two, is rather high.

Fuzzy set theory produces the following result:

[3]Note that the same rules for intersection and conjunction hold in fuzzy set theory as in classical set theory.

$$\begin{aligned} c_{Chair \cup Cushion}(Beanbag) &= max(c_{Chair}(Beanbag); c_{Cushion}(Beanbag)) \\ &= max(0.8; 0.5) \\ &= 0.8 \end{aligned} \tag{2.2}$$

As you can see, this works really well.

However, some problems arise, such as problems concerning the conjunction fallacy. These problems are pointed out by Osherson and Smith in their article ‘On the adequacy of prototype theory as a theory of concepts’ [Osherson and Smith, 1981]. This article will be discussed below.

2.3 Osherson and Smith’s criticism of Fuzzy Theoretic Combinations

Osherson and Smith clearly show that, even though the fuzzy set theory in combination with prototype theory looks fairly promising, there are several flaws.

2.3.1 Fuzzy set theory and the universally true and false sentences

First, fuzzy set theory gives some really strange results when confronted with sentences that should be always either false or true. For example the sentence “A beanbag is a Cushion or is not a Cushion” should always be true (for it has the logical form of $p \vee \neg p$, which always results in the truth-value 1). However, this is not the case.

As we stated earlier, a beanbag has a graded membership to the ‘cushion class’ of 0.5: $c_{Cushion}(Beanbag) = 0.5$. Therefore, the graded membership of the ‘non-cushion class’ should also be 0.5: $c_{nonCushion}(Beanbag) = 0.5$, because $c_{nonA}(x) = 1 - c_A(x)$. But now some problems arise, because the sentence “A beanbag is a Cushion or is not a Cushion” will be translated as:

$$
\begin{aligned}
c_{Cushion \cup nonCushion}(Beanbag) &= max(c_{Cushion}(Beanbag); c_{nonCushion}(Beanbag)) \\
&= max(c_{Cushion}(Beanbag); 1 - c_{Cushion}(Beanbag)) \\
&= max(0.5; 0.5) \\
&= 0.5 \qquad (2.3)
\end{aligned}
$$

It is obvious that this answer is false, because the statement ‘A beanbag is a cushion or is not a cushion’ should always be true, and therefore the characteristic function should always assign a 1 to this sentence, which it does not.

The same –obviously– happens when we take the union of Cushion and non-Cushion. This should always result in a 0 (because being a Cushion and not being a Cushion at the same time is a contradiction), but receives by the deduction shown below, the same value as the previous sentence, 0.5.

$$
\begin{aligned}
c_{Cushion \cap nonCushion}(Beanbag) &= min(c_{Cushion}(Beanbag); c_{nonCushion}(Beanbag)) \\
&= min(c_{Cushion}(Beanbag); 1 - c_{Cushion}(Beanbag)) \\
&= min(0.5; 0.5) \\
&= 0.5 \qquad (2.4)
\end{aligned}
$$

These were examples where fuzzy set theory leads to logical contradictions, when evaluating sentences that are universally true or universally false.
Next, we are going to look at the way fuzzy set theory handles the conjunction fallacy.

2.3.2 Fuzzy set theory and the Conjunction Fallacy

Second, Osherson and Smith consider examples where certain objects seem to be better instances of a conjoined concept than of its elementary constituent concepts[4]. This occurrence of an object that has a higher probability of being a conjunct of two things than being just one of those things, is

[4]There are many more examples, like the cigarette tax-example, the Italian Rail-example [Tentori et al., 2004] , the Tom-example and the Linda-example [Kahneman, 2002]

called the conjunction fallacy. It was first reported by Tversky & Kahneman [Tversky and Kahneman, 1983].

Example 2.2

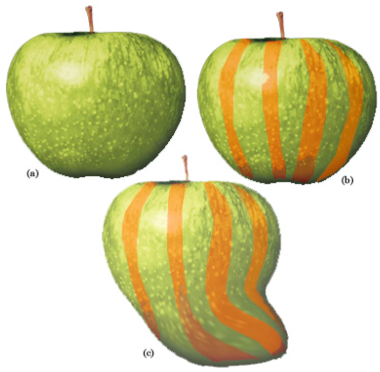


Figure 2.2: Three apple-like objects

Consider the three objects of figure 2.2. Two of them are striped, the third is not. As one can see, the object (c) is not a prototypical apple when it comes to shape, but the objects (a) and (b) are quite prototypical.
Let us now consider apple (b). It is obvious that, although it has the right shape, this is not a very prototypical apple. When we look in our fruit bowl, most apples look like (a), because apples are normally not striped.
However, exemplar (b) is a fairly good prototype of the conjunct 'striped apple'.

We can formulate this empirical observation in the following way:

$$c_{StripedApple}(b) > c_{Apple}(b) \tag{2.5}$$

Because a striped apple can be seen as a combination of 'striped' and 'apple', it is a combination of the two concepts. This is called 'concept combination' in prototype theory, and it leads us to the following inequality:

$$c_{Striped \cap Apple}(b) > c_{Apple}(b) \tag{2.6}$$

However, in fuzzy set theory this leads to a problem, since the conjunction should receive the lowest value of the two conjuncts:

$$c_{Striped \cap Apple}(b) = min(c_{Striped}(b); c_{Apple}(b)) \leq c_{Apple}(b) \qquad (2.7)$$

This is a contradiction with (2.5)! Hence, fuzzy set theory produces a counterintuitive result in the present case.

This contradiction occurs because of monotonicity. Everything that belongs to the set $a \wedge b$ will also belong to the set a ($a \cap b \subseteq a$), but not necessarily the other way around. Therefore, due to monotonicity, it will always be the case that $P(a) \geq P(a \wedge b)$ and therefore, (2.7) leads to a contradiction.

2.4 Conclusion

In this chapter, we evaluated a method defined by Osherson and Smith in their article 'On the adequacy of prototype theory as a theory of concepts' [Osherson and Smith, 1981]. They tried to implement prototype theory with fuzzy set theory, but stumbled upon two major problems.

Osherson and Smith themselves conclude that prototype theory combined with fuzzy set theory cannot be the right way to go, because they cannot solve the conjunction fallacy and the definite values for universally true and universally false sentences. However, they do not prove that the idea of prototype theory itself is wrong. Merely the implementation through fuzzy set theory is.

This has led to an article of Hans Kamp and Barbara Partee [Kamp and Partee, 1995], who agree with Osherson and Smith that fuzzy set theory is not the right way to implement prototype theory. Instead, they present a new theory, supervaluation theory, as a better way to do this.
They claim that their way of implementing prototype theory can handle even the conjunction fallacy. Therefore, I will discuss their article in the next chapter.

Chapter 3

Prototype Theory and Supervaluation Theory

Kamp and Partee extend Osherson and Smith's method of using the characteristic function to determine whether an object is prototypical for a certain class or not. But instead of one characteristic function, Kamp and Partee introduce two different functions. They make a difference between c^e, which denotes the degree of membership of a certain object to a certain class, and c^p, which denotes the degree of closeness to the prototype of the concept. Note that the c-function used by Osherson and Smith is equivalent to Kamp and Partee's c^p function.

This division solves the problem that we addressed earlier: a whale is not very close to the prototypical mammal, because it swims in the see and it looks like a fish. Therefore, the c^p function will probably assign a rather low value to the whale when it comes to determining its closeness to the prototypical mammal. On the other hand, c^e will probably assign a rather high value to the whale for being a member of the class of mammals, because it undoubtedly is a member.
The same will hold for penguins. They are certainly not prototypical birds (so their c^p value will be rather low), but they are definitely birds (and their c^e value will therefore be accordingly high).

3.1 A bit more on c^p

At this point, we are going to take a little time out to take a closer look at the function c^p, because the notion of c^e is rather clear but there has been a lot of debate about the definition of c^p.

Eleanor Rosch [Rosch, 1999] conducted an experiment, where subjects had to categorize objects into classes. She used three levels of classes:

Super ordinate classes like 'furniture' and 'tree'

Basic classes like 'chair' and 'table'

Subordinate classes like 'kitchen chair' and 'dining-room table'

She showed that the subjects classified significantly more objects into the Basic and the Subordinate level classes than in the Super ordinate level classes.

Armstrong, Gleitman and Gleitman [Armstrong et al., 1983] conducted a second experiment, where subjects had to tell whether instances were prototypical of a certain class. Half of the classes they used, were –just as with Rosch– the so called prototypical classes, such as 'sport', 'vegetable' and 'vehicle'. The other half of the classes were well defined classes like 'odd number' and 'female'.

In the first category, it is perfectly acceptable to find a graded membership, but it would be rather odd to find such a thing in the second category, because a number is definitely even or odd, but not something in between. Even stronger: there are no objects that are on the boundary of the odd or even class: a number is always definitely odd or definitely even.

The students in this experiment were asked to rate the extent to which each instance represented their idea of the meaning of each class on a 7-point scale.
The results obtained from the 'prototypical classes' were essentially identical with the results obtained by Rosch. The striking result however, was that the subjects also gave the instances from the well-defined classes a graded membership. For example, the subjects thought that 3 was a better odd number than 501 and 'mother' was a better female than a 'comedienne'.
These outcomes can be explained by interpreting the c^p measured here, as a typicality measure. Therefore, 3 is a better odd number than 501 because 3 is the first number that comes to mind when we are asked to name a prime number.
For more information on this research, see the full paper [Armstrong et al., 1983].

However, this interpretation of c^p is not the only possible interpretation. In his Nobel lecture [Kahneman, 2002], Daniel Kahneman discusses many different problems with typicality. Among them is his famous Linda example [Tversky and Kahneman, 1983]:

Linda is 31 years old, single outspoken and very bright. She majored in philosophy. As a student she was deeply concerned with issues of discrimination and social justice and also participated in antinuclear demonstrations.

Subjects had to grade her on two items: 'Linda is a bank teller' and 'Linda is a bank teller and active in the feminist movement'.
The outcome was that Linda resembles the image of a feminist bank teller more than she resembles the image of a stereotypical bank teller. This outcome would suggest that the c^p measured here, is a probability and not, as it was in Armstrong's paper, a typicality measure that is calculated by the following procedure: good birds are these exemplars that come to mind first.

These two interpretations of c^p are not necessarily conflicting. But for clarity purposes, we will hold on to the definition of Kahneman in this chapter.

3.2 Supervaluation of Kamp and Partee

As has been pointed out in the previous chapter, Osherson and Smith clearly show that there are some problems with using fuzzy set theory to implement prototype theory. One of the problems was that the universally true and universally false sentences did not receive the values they should receive in fuzzy logic (respectively 1 and 0). Kamp and Partee show that this problem can be solved by using (an extended version of) the supervaluation method.

3.2.1 Supervaluation

Let us first take a closer look at the supervaluation method itself. The method Kamp and Partee present is based on the supervaluation theory of Bas van Fraassen[1], that was presented in his article 'Singular terms, Truth-

[1]Note that the supervaluation method presented by Kamp and Partee in their article, is actually an extension of the supervaluation method presented by Van Fraassen. Kamp & Partee have extended the supervaluation method with a measure function, which will be discussed later.
Note that from now on, always when the supervaluation theory is mentioned, the extended version is meant.

value gaps, and Free Logic' [Fraassen, 1966][2]. In this method, a partial function Π assigns a value for the degree of membership (c^e) to an object, that expresses whether this object is a member of the concept class or not. The possible values are 0, 1, but it can also leave a value 'undefined'. Note that 'undefined' is not the same as 'intermediate' or 0.5, it just means that there is not yet a definite value for this object. This supervaluation method is therefore a two-valued logic, and not a three-valued one.
Note that the logic described here is a propositional logic. In this logic, we use a set of atoms:

$$\text{ATOMS: } \{p_1, \ldots, p_n\} \tag{3.1}$$

Each atom is a well formed formula. Furthermore, if ϕ and ψ are well formed formulas, then so are $\neg\phi$, $\phi \wedge \psi$ and $\phi \vee \psi$.

In order to determine whether these formulas are true or false (respectively 1 or 0) , we need a valuation function v, such that it assigns to every atom in the language a value:

$$v: \text{ ATOMS } \Rightarrow \{1, 0\} \tag{3.2}$$

This valuation function is a *complete* function. This can also be described in set theoretic terms:

Definition 3.1. *A function f is complete, if for all objects x from set A, there exists exactly one object y in set B, such that (x, y) is in f.*
$\forall x \in A, \exists!\ y \in B$: $(x, y) \in f$
$f \subseteq A \times B$

We also need a partial valuation function Π, in order to be able to give every well formed formula in the language a valuation, or to leave the valuation 'undefined'. For example, if we consider the apples of figure 2.2 and we want to know whether the objects are apples or not, the partial function will give a value to (a) and (b), but not to (c), because it is unclear whether or not it is an apple.

[2]Note that van Fraassen uses a different notation than Kamp & Partee. For reasons of clarity, I will only use Kamp & Partee's notation.

Therefore, the partial valuation function Π is a subset of the valuation function v, because (in the present case) the valuation function will give a value to all three objects, where Π leaves the value for (c) undefined:

Definition 3.2. *A partial valuation function* Π *is a valuation function, such that*
$\Pi \subseteq v$

To be able to represent all of the methods that Kamp and Partee use in a propositional logic, we make a small change to the original language. Now, instead of simple atoms, we use propositional symbols:

$$\text{PROP: } \{P(a_1), \ldots, P(a_n)\} \tag{3.3}$$

This way, supervaluation theory has the effect that if a statement ϕ has no truth-value (is 'undefined'), then so do for example $\phi \wedge \phi$ and $\phi \wedge \psi$. But it still holds that the universally false sentence (e.g. $\phi \wedge \neg\phi$) is always false and the universally true sentence (e.g. $\phi \vee \neg\phi$) is always true.

Where the partial function Π gives us a normal valuation, the complete function v gives us a supervaluation, in order to give the undefined sentences a truth-value as well. This function looks at all the possible completions, and tries to find out whether these completions can fill the truth-value gap that was created by the undefined objects.

This results in the following:
If the partial valuation function Π assigns the value 1 to $P(x)$, the object is an element of the positive extension of the model. If zero is assigned to the object x, than the object is an element of the negative extension of the model. If the value for the object is neither very high, nor very low, x will be assigned a supervaluation to determine whether or not it belongs to the class P.

In other words:

Definition 3.3.

$$\begin{aligned} x \in ||P||^+_\Pi &\qquad \textit{if } \Pi(P(x)) = 1 \\ x \in ||P||^-_\Pi &\qquad \textit{if } \Pi(P(x)) = 0 \\ x \in v - (||P||^+_\Pi \cup ||P||^-_\Pi) &\qquad \textit{else.} \end{aligned} \tag{3.4}$$

When using supervaluation, we obtain the following:

Definition 3.4. $\Pi^\star$ *is the set of all completions* Π' *of* Π.

For $\Pi^\star$ it holds that:

Definition 3.5. *The truth-value of a sentence* ϕ *with respect to* $\Pi^\star$ *is:*

- *1 if its truth-value is 1 in all completions* $\Pi' \in \Pi^\star$
- *0 if its truth-value is 0 in all completions* $\Pi' \in \Pi^\star$
- *undefined otherwise (e.g., if 0 in some and 1 in others)*

To explain in more detail how supervaluation theory works, we will take a look at an example borrowed from Labov [Labov, 1973].

Example 3.1
In his article ('Boundaries of words and their meanings', [Labov, 1973]), Labov investigates what the boundaries of a concept, in this case the concept 'cup', are. Some cups are more like (wine)glasses and others more like bowls and some are quite prototypical cups.
He presented his cups to a group of people in different contexts. This test group had to say whether or not the presented object was a cup.

In the end, the result was that whether something was considered a cup or not, depended on the ratio height-width and whether or not the object had a handle.

Let us now use these results for an example to show how supervaluation works. In the figure below, 4 cups are represented. The cup shape (the width-height ratio) is evaluated in the middle column. Furthermore, we have a column for the greyness. This will be used in the second part of the example.

They all have handles, so the defining attribute for whether these objects are considered cups or not, is the width/height ratio.

The first thing the supervaluation method of Kamp and Partee does, is calculating the degree of membership (c^e) of every object, for every attribute.

(a) cup1 (b) cup2 (c) cup3 (d) cup4

Figure 3.1: Four different cups.

Object	cup shape	greyness
cup1	Perfect cup shape	Not very grey
cup2	A bit too wide for its height	Quite grey
cup3	Too wide for its height	Grey
cup4	Far too wide for its height	Almost grey

Figure 3.2: Four different cups, how cup shaped and how grey they are.

The resulting values for the degree of membership given by the partial function Π are the following:

$$\begin{aligned} c^e_{cup}(cup1) &= 1 \\ c^e_{cup}(cup2) &= 0.75 \\ c^e_{cup}(cup3) &= 0.5 \\ c^e_{cup}(cup4) &= 0.25 \end{aligned} \tag{3.5}$$

This means that the partial function Π will assign a 1 to cup 1 (because it has the perfect shape) and 'undefined' to all the other cups. In order to give all the other cups a truth-value as well (give them a 'supervaluation'), it is necessary to look at all the completions Π'. These different completions Π' will divide up the original range of indetermination in different ways.
Because cup2 is 'closer' to the positive extension of 'cup' than cup3, it is reasonable to suppose that this comparison is reflected by the set of completions $\Pi^{\star}$. In other words, the set where cup3 belongs to the positive extension of 'cup', must be a proper subset of the set where cup2 belongs to the positive extension of 'cup'.
This means that the set of completions in which an object belongs to the extension of a certain concept indicates the degree to which that object falls under that concept.

Each completion Π' has a measure function μ that assigns a value between zero and one, [0,1], to a "sufficiently rich subfield of the set f completions of $\Pi^{\star}$" [Kamp and Partee, 1995]. These will from now on be represented as triples $< \Pi, S, \mu >$, where Π is the partial function, S is a set of completions and μ is the measure function.

Such a triple $< \Pi, S, \mu >$ defines for each concept a, a fuzzy characteristic function A. This function is defined by [Kamp and Partee, 1995]:

Definition 3.6. *For any $a \in v$, $\mu^e_A(a) = \mu(\Pi' \in S:$ a is in the extension of A in $\Pi')$ where v' is the set of valuations which eliminate all truth-value gaps by extending the positive and negative extension of each predicate so that they jointly exhaust the domain.*

Note that μ^e_A can be identified with c^e_A, but not with c^p_A, because when $\mu_A(a)$ is in the positive extension, $\mu_A(a)$ will always be 1 and when it is in the negative extension, it will always be 0. Only when the object is in the truth-value gap, the measure function μ can assign a value between 0 and 1. Therefore, this function can not be identified with the measure of prototypicality (which can be graded) but it can be identified with the measure of membership[3].

In the following figure, the results are shown. We have four possible completions and four μ-values. In this case, all the μ-values are 0.25, because of definition 3.6.

	C(cup1)	C(cup2)	C(cup3)	C(cup4)	μ
Π	1				
Π'_1		0	0	0	0.25
Π'_2		1	0	0	0.25
Π'_3		1	1	0	0.25
Π'_4		1	1	1	0.25

Figure 3.3: Valuation function Π, its four possible completions Π' and their μ-values

Now, we'll do the same with another attribute: the colour. If we want to find a 'grey cup', we will need to take the intersection of the two attributes in order to find an appropriate example of a grey cup.
The colour's c^e functions will be as follows:

[3] Recall the experiment of Armstrong, Gleitman and Gleitman [Armstrong et al., 1983], that was discussed above in the 'A bit more on c^p'-section.

$$\begin{aligned} c^e_{grey}(cup1) &= 0.5 \\ c^e_{grey}(cup2) &= 0.75 \\ c^e_{grey}(cup3) &= 1 \\ c^e_{grey}(cup4) &= 0.9 \end{aligned} \tag{3.6}$$

The value for cup3 is a one, but cup1, cup2, and cup4 will receive an intermediate value. The values given by the partial function A' result in the following completions Π' and the μ-values are calculated in the same way as before.

	G(cup1)	G(cup2)	G(cup3)	G(cup4)	μ
Π			1		
Π'_1	0	0		0	0.1
Π'_2	0	0		1	0.15
Π'_3	0	1		1	0.25
Π'_4	1	1		1	0.5

Figure 3.4: Valuation function Π, its four possible completions Π', and their μ-values

These possible completions Π' for both cup shape and greyness form together the set of completions $\Pi^\star$ of the conjunct. The 'super μ-values' are obtained by multiplying the μ-values of the possible completions. This is allowed according to the Kolmogorov axiom, because the two attributes (cup shape and greyness) are completely independent.

Definition 3.7. *$\mu_{A\cap B}$ is equivalent to $\mu_A \times \mu_B$ iff A and B are independent.*

The results of this are presented in figure 3.5.

3.2.2 Supervaluation and the universally true and false sentences

Now let us return to the problems at hand, the problems Osherson and Smith faced when implementing Prototype theory with fuzzy set theory.

	C(cup2)	C(cup3)	C(cup4)	G(cup1)	G(cup2)	G(cup4)	μ
Π'_1	0	0	0	0	0	0	0.025
Π'_2	0	0	0	0	0	1	0.0375
Π'_3	0	0	0	0	1	1	0.0625
Π'_4	0	0	0	1	1	1	0.125
Π'_5	1	0	0	0	0	0	0.025
Π'_6	1	0	0	0	0	1	0.0375
Π'_7	1	0	0	0	1	1	0.0625
Π'_8	1	0	0	1	1	1	0.125
Π'_9	1	1	0	0	0	0	0.025
Π'_{10}	1	1	0	0	0	1	0.0375
Π'_{11}	1	1	0	0	1	1	0.0625
Π'_{12}	1	1	0	1	1	1	0.125
Π'_{13}	1	1	1	0	0	0	0.025
Π'_{14}	1	1	1	0	0	1	0.0375
Π'_{15}	1	1	1	0	1	1	0.0625
Π'_{16}	1	1	1	1	1	1	0.125

Figure 3.5: All possible completions for cup shape and greyness, with their μ-values.

Recall the problems fuzzy set theory ran into when it had to handle sentences like ‘A beanbag is a cushion and is not a cushion’ or ‘A beanbag is a cushion or is not a cushion’, with $c_{Cushion}(Beanbag) = 0.5$ and therefore $c_{nonCushion}(Beanbag) = 0.5$ as well.

The following happens when we use Kamp and Partee’s supervaluation theory:
Because $c_{Cushion}(Beanbag) = 0.5$ (neither a very high, nor a very low value), $Cushion(Beanbag)$ receives no truth-value. Now Cushion(Beanbag) will receive a definite truth-value in each completion Π', 0 in some, 1 in others. Whichever value it will receive, the negation will always be 1 minus the assigned value.

The resulting $\Pi^\star$ is represented in figure 3.6.

	$Cushion(Beanbag)$	$\neg Cushion(Beanbag)$
Π'_1	1	0
Π'_2	0	1

Figure 3.6: Possible completions for Cushion(Beanbag)

Now, as you can see in figure 3.6 in every completion Π' the sentence 'A beanbag is a cushion or not a cushion' will receive a 1 and the sentence 'A beanbag is a cushion and not a cushion' will receive a 0, by using the same rules for concept combination as Smith and Osherson used.[4] This clearly shows that the first problem of fuzzy logic is solved by using supervaluation theory. But does supervaluation theory solve all the problems of fuzzy logic? That is what we will try to find out next.

3.2.3 Supervaluation and the Conjunction Fallacy

The other major problem that fuzzy logic could not handle was the conjunction fallacy illustrated by the problem of the striped apple.
Kamp and Partee admit that this problem is still present when using supervaluation theory, because in their theory it is still impossible for a conjunction to receive a higher (or equal) value than each of its conjuncts. They also admit that this problem cannot be solved with a different interpretation of the c-function, because the problem still exists for both c^e and c^p.

Kamp and Partee acknowledge the shortcomings of supervaluation, because they define c as a probability measure. Therefore, c inherits all the properties of probabilities. One of those properties is that a conjunct of two probabilities can never be higher than a single conjunct.

3.3 Supervaluation Method with Recalibration

In order to make their supervaluation method work in the general case, Kamp & Partee introduce a patch: recalibration. They also redefine the problems they had with the conjunction fallacy. Whether these patches are actually good solutions, will be discussed in this section.

Kamp and Partee claim that the main problem with the striped apple is not a logical combination problem, but a semantic problem. They claim that the rule that was presented in the beginning of their article, that a combination of two concepts could be represented with the intersection of the two concepts, might be wrong. As an example, they give the following three sentences:

[4] Recall that concept combination was implemented in fuzzy logic as an intersection of two sets:
$c_{A\cap B}(x) = min(c_A(x); c_B(x))$

i Sam is a giant and a midget

ii Sam is a giant midget

iii Sam is a midget giant

Here, it is obvious that the concept 'giant midget' cannot be formed with the intersection of 'giant' and 'midget'. All three sentences have a different meaning, but both supervaluation theory and fuzzy logic would treat them as one and the same sentence. In order to overcome this problem, Kamp and Partee propose to use recalibration. These recalibrations are involved in determining the extensions of the compounds of the concept combination. For the term 'striped apple' this means to determine the value of the term, we need to go through a two-stage process:

1. All the best cases of 'striped' within $||apple||_\Pi$ are treated as definitely within the positive extension of $striped|apple$ ("striped relative to apple") and the worst cases of striped within $||apple||_\Pi$ as definitely in the negative extension of $striped|apple$. The intermediate cases are adjusted proportionally.
2. Now the supervaluation theory is used to determine the value of the object for the combination 'striped apple'

They use the following formula for this recalibration:

$$c^P_{Striped|Apple}(a) = \frac{c^P_{Striped}(a) - c^-_{Striped|Apple}}{c^+_{Striped|Apple} - c^-_{Striped|Apple}} \tag{3.7}$$

Where

$$c^+_{Striped|Apple} = sup\{c_A(a) : a \in ||N||_\Pi\} \tag{3.8}$$

and

$$c^-_{Striped|Apple} = inf\{c_A(a) : a \in ||N||_\Pi\} \tag{3.9}$$

Now, the supervaluation theory is used, just as before, to calculate the degree to which any object a satisfies the conjunction $StripedApple$. Only this time, not the original characteristic function c is used, but rather a new function c' where c' already incorporates the recalibration of $Striped$ in the context of *Apple*.

Let us again consider the three apples of figure 2.2 and let us assume that the values for $c^p_{Striped}$ are the following:

$$\begin{aligned} c^p_{Striped}(a) &= 0.1 \\ c^p_{Striped}(b) &= 0.9 \\ c^p_{Striped}(c) &= 0.8 \end{aligned} \tag{3.10}$$

And consequently, according to (12) and (13),

$$\begin{aligned} c^+_{Striped|Apple} &= 0.9 \\ c^-_{Striped|Apple} &= 0.1 \end{aligned} \tag{3.11}$$

.

Now, we can calculate how good a prototype for example apple (c) is:

$$\begin{aligned} c^P_{Striped|Apple}(c) &= \frac{c^P_{Striped}(a) - c^-_{Striped|Apple}}{c^+_{Striped|Apple} - c^-_{Striped|Apple}} \\ &= \frac{0.8 - 0.1}{0.9 - 0.1} \\ &= 0.875 \end{aligned} \tag{3.12}$$

This clearly tells us that (c) is a pretty good example of a striped apple.

3.4 Discussion

Looking at all the material that Kamp and Partee presented in their article, we have found three major flaws in their theory. In this section these flaws will be discussed.

First, as Jules van Ligtenberg shows in his Bachelor Thesis [Ligtenberg, 2007], recalibration has a flaw concerning the c^p function. Kamp and Partee do not clearly show which c-function needs recalibration. They assume that if c^p needs recalibration, only the positive extension needs to be recalibrated. In general, this will work fine, as long as the objects we want to combine are natural objects. But Osherson and Smith have already shown that in the case of the striped apple, many subjects did not imagine a natural apple, when asked to imagine a striped apple. Instead, they imagined something striped, with the outline of an apple. That way, the resulting object will definitely have a higher value for the combination than for one of the constituents.

Second, if the value of $c_{StripedApple}$ was intermediate before recalibration, it will now still be intermediate after recalibration. In other words: the recalibration has no effect at all. This is of course a huge problem.

Third, Kamp and Partee seem to present some nice solutions, like the new interpretation of concepts, but not general ones. For example, it does not work for the examples like 'stone lion', because in this case, it cannot adjust to the fact that in this case, the end product of the combination AN, has all the properties of the adjective A, but only the shape of the noun N. This is exactly the other way around from the striped apple case. Therefore, the words in the combination must be reinterpreted (e.g. 'stone' means actually 'made of stone').
This makes their theorem highly context dependent, because for every combination of two concepts, it might be necessary to reinterpret the concepts. But Kamp & Partee have found no good solution for this reinterpretation yet, they only give a rather tentative sketch of the solution.

3.5 Conclusion

In this chapter, I discussed the implementation of Prototype theory by using supervaluation, as was presented in the article 'Prototype theory and compositionality' by Kamp and Partee [Kamp and Partee, 1995] and the way the conjunction fallacy is handled by supervaluation.
It turned out that Kamp and Partee do not give a general solution for the conjunction fallacy. Their theory can handle the universally true and false sentences, but cannot handle the conjunction fallacy.
Although supervaluation combined with recalibration can handle the striped apple case, it is still too specific to give a nice solution for the general case.

Chapter 4

Prototype Theory and Default Logic

In the previous chapters, we discussed the implementation of Prototype Theory and the handling of the conjunction fallacy by fuzzy sets and supervaluation theory. This chapter will discuss yet another way of implementing Prototype Theory: Default Logic.

4.1 Default Logic

Default logic is a non-monotonic logic. It was first presented by Raymond Reiter in 1980 [Reiter, 1980] and developed to manage default assumptions that human beings handle in every day life.
A non-monotonic logic is a logic that ensures that if one can draw conclusion ψ from the premises $\phi_1, \ldots, \phi_n$, it does not automatically hold that one can draw the same conclusion ψ from the premises $\phi_1, \ldots, \phi_{n+1}$.
This means that default logic uses the closed world assumption: as long as there is no proof of the opposite, we conclude that the default rule is correct[1].

For example, the next reasoning is valid[2]:

[1]Note that this closed world reasoning is closely related to bounded rationality, as presented by Daniel Kahneman in his Nobel prize lecture [Kahneman, 2002]. Bounded rationality states that human beings are bounded in their reasoning by time and possible knowledge. For example, it is not possible to check every statement against all knowledge available in the world, for there is just too little time to do so and there is so much knowledge available, that it is impossible to cover it all. It would be an endless exercise. Therefore, there is some evidence that the closed world assumption is an assumption that human beings do actually use (although not deliberately).

[2]NB The interpreter of the next sentences should not know that ostriches cannot fly

Prerequisite 1: Birds can typically fly. (4.1)
Prerequisite 2: An ostrich is a bird.
Prerequisite 3: Oz is an ostrich.
Conclusion: Oz can fly.

But the conclusion 'Oz can fly' will be invalid as soon as we gather a new bit of information that tells us that ostriches cannot fly. Therefore, the next reasoning is invalid:

Prerequisite 1: Birds can typically fly. (4.2)
Prerequisite 2: An ostrich is a bird.
Prerequisite 3: Oz is an ostrich.
Prerequisite 4: Ostriches cannot fly.
Conclusion: Oz can fly.

In this case, adding an extra prerequisite changes the conclusion radically: it swaps the positive conclusion for the complete opposite. This is called non-monotonicity.

Note that the first sentence of both reasonings is 'Birds can *typically* fly'. The word 'typically' is often used in default logic. It describes the so-called default rule. This means that as long as we have no information that falsifies it, we that for example whenever we see a bird, we assume that it can fly.
Note that this usage of default rules implies that default logic uses the notion of c^p presented by Kahneman and Tversky [Tversky and Kahneman, 1983]. Default logic is basically a formalization of the typicality measure, where the default rule states what is typically an attribute of a certain class.

4.2 Veltman's Default Logic and the Conjunction Fallacy

In his article "Een zogenaamde denkfout" [Veltman, 1998], Veltman shows that, when using default logic, the conjunction fallacy is no fallacy at all.

First, a set of rules is introduced, which should be able to handle prototypicality.
He introduces a null-state **0**. An agent is in this state, if he has had no information at all.
Let us now assume that we are in state s, if we receive some information ϕ, the state s will be updated.

Definition 4.1. *The state s will be updated to state s' when the agent in state s receives some additional information ϕ:*
$s' = s + \phi$

Note that the +-sign in the definition is not the normal mathematical +-sign. For example, $s + \phi + \psi$ is not equal to $s + \psi + \phi$, where in mathematics it is[3].

A conclusion ψ is only valid if and only if the conclusion follows directly from the premises. In other words: a conclusion is only valid when the state does not change after the conclusion is drawn.

Definition 4.2. *A conclusion ψ is valid iff:*
$s + \mathbf{0} + \phi_1 + \ldots + \phi_n + \psi = s + \mathbf{0} + \phi_1 + \ldots + \phi_n$

Now that we have the rules of this language, we can investigate the effects of using this particular logic on the conjunction fallacy.

4.2.1 Default logic and the universally true and universally false sentences

In the previous chapters, we saw that universally true and universally false sentences could not be handled by fuzzy set theory, but were correctly handled by the supervaluation theory. Now, the question is: are they handled correctly by default logic as well?

Sentences like: *A beanbag is a chair or a cushion* should always be validated as true. Does default logic do this correctly?

[3]For example, the sequence *There is someone at the door ... It's probably Nicholas ... It's Pete.* is consistent, where the following sequence is not: *There is someone at the door ... It's Pete ... It's probably Nicholas.* (Free translation of example in [Veltman, 1998]).

First of all, remember that the (graded) membership value of Chair(Beanbag) was 0.8 and of Cushion(Beanbag) 0.5. Note that when translating[4] such values into sentences, we use the word 'normally' only when the membership value is bigger than 0.5. Because the word 'normally' indicates that the sentence is a default rule. As soon as the membership value is lower than 0.5, the sentence can be translated into a negative default rule (like *A beanbag isn't normally not a chair*). However, we also have here the case where the membership value is precisely 0.5. This cannot be translated into a default rule, therefore, we do not use the word 'normally' here.[5].
This means that we can translate these values into:

- A beanbag is normally a chair.
- A beanbag is sometimes a cushion.

Because we are using the closed world assumption in default logic, and we have not found any clue that would suggest that a beanbag is something different than a chair or a cushion, we can now conclude that:
A beanbag is normally a chair or a cushion.

Note that this is only possible, because a beanbag has many properties that are normal for a chair or a cushion (such as usage, shape, etc), but are not normal for objects that are not chairs or cushions.

The same holds for the universally false sentences, like: *A beanbag is a chair and a cushion.*

From the prerequisites mentioned above it is not possible anymore to conclude that the sentence *A beanbag is normally a chair and a cushion*, because the second premise gives us the evidence that sometimes a beanbag is not a cushion! Because this second sentence is no default rule (because it has only the membership value of 0.5), we cannot conclude that all beanbags are normally chairs and cushions.

So, default logic handles the universally false and true sentences correctly. Let us take a look at how default logic handles the conjunction fallacy.

[4] This is important: Veltman creates sequences of representativeness, such as 'A salmon is a more prototypical fish than a goldfish', but does not use a probabilistic measure. Therefore, this translation is needed

[5] Note that these translations are not formalized, there are no rules for how to translate membership values into natural language sentences. Therefore, please keep in mind that this is just my interpretation of the translation

4.2.2 Default logic and the Conjunction Fallacy

In the previous chapters, we have considered theories that behave in a monotonic way. This 'created' the conjunction fallacy. Default logic is a non-monotonic logic. So, our first impression is that it should be able to handle the conjunction fallacy. In this section, I will investigate whether this impression is correct.

Let us again consider the three apples of figure 2.2. Now we can state that *Apples normally look like (a)*, because apple (a) has a perfect apple-shape and colour.
We can also say that *Stripes normally look like the ones on (b)*.
Now, we know that (a) is a better example of an apple than (b), because, although (a) and (b) have the same shape, (a) has a better 'apple-colour' than (b). We also know that (b) is a much better example of a striped object than (a), because (a) has no stripes at all.
We also know that (b) is a better example of an apple than (c).

Therefore, we have the premises:

- (a) is a better example of an apple than (b)
- (b) is a much better example of a striped object than (a)

From these premises we can conclude that (b) is a better example of a striped apple than (a), because (b) has a property (the stripes!) that is unusual for apples, but common for striped apples, where (a) has a property (no stripes) that is common for apples, but unusual for striped apples.
In the previous chapters, this would have led to a contradiction, but in default logic, it is a valid conclusion. Therefore, the conjunction fallacy is no fallacy anymore, when using default logic.

4.3 Conclusion

Although this way of handling the conjunction fallacy and the universally true and false sentences looks very promising, it is not a complete solution. In the previous chapters we made it very clear to what degree an object was a member of a certain class. When using default logic, this is less clear, because there is no probabilistic measure that can be used. A sentence like: *Normally, birds can fly* does not indicate in how many cases a bird can actually fly.

On the other hand, default logic may be a more natural way of handling human reasoning. Because there is some evidence that human beings use prototypicality rather than probability and that human beings are bounded in their reasoning and the closed world assumption is a rather good implementation of this bounded rationality, it would not be very farfetched to conclude that although default logic does not always lead to the right answers, it does lead to the most 'human-like' answers.

In the next part, we will take a look at a different approach to concept combination and concept representation altogether. We will investigate whether these approaches can handle the conjunction fallacy as gracious as default logic can and will try to find out if there is an overlapping idea at the basis of the two approaches.

Part II

Concept combination using a non-Boolean algebra

Chapter 5

Using Geometrical models in Concept Combination

In the previous chapters we discussed the different implementations of the prototype theory. We examined three different options (fuzzy logic, (an extended form of) supervaluation theory and default logic). What the first two options have in common, is that they use a characteristic function to determine how close an object is from the prototype of the concept or what the degree of membership from the object to a certain class is and both options are based on Boolean statistics.
In the next few chapters we will take a closer look at options that are based on a non-Boolean algebra.

5.1 Dividing the concept space with Voronoi tessellation

Gärdenfors presents in his book 'Conceptual Spaces: the geometry of thought' [Gärdenfors, 1995] a slightly different approach. In order to determine to which class an object belongs, he divides the object space using the Voronoi tessellation. Now, the object space is divided in different areas that have (in contrast to the Osherson method) no overlap[1]. In other words, the classes have sharp and not vague boundaries. The difference between the two systems is explained in figure 5.1. On the left is 'Osherson's method' and on the right is 'Gärdenfors's method' using Voronoi tessellation.

The Voronoi tessellation draws boundaries exactly between the prototypes and creates, by doing this, a convex geometrical space.

[1]Therefore, graded membership (a value between 0 and 1 for c^e) is no longer possible

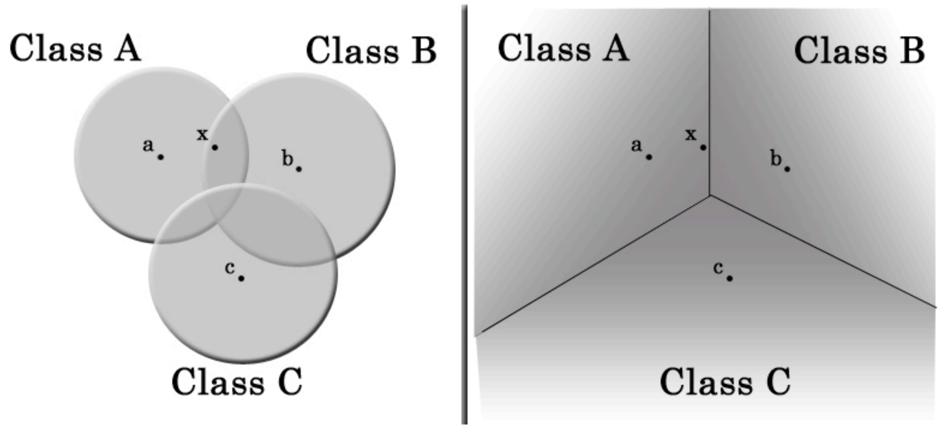


Figure 5.1: Two ways of class representations. Fuzzy sets on the left, the result of the Voronoi tessellation on the right.

In figure 5.1 we show the different classes as two dimensional representations. But with the division of the space by the Voronoi tessellation, it is also possible to represent multidimensional concepts, such as colour. This has the effect that it is possible to make a multidimensional space that represents a concept. With this multidimensional space, it can be possible to make concept combinations in a more natural –and non-Boolean– way.

This is what Peter Gärdenfors did in several articles and a book [Gärdenfors, 1993, 1995, 1998, 2000]. These will be discussed next.

5.2 Gärdenfors – The Geometrical model

One big difference between the approaches of Osherson & Smith and Kamp & Partee on the one side and Gärdenfors on the other side, is the psychological foundation of their approaches. As explained earlier, Osherson & Smith and Kamp & Partee use the prototype based cognition as a basis of their approach[2]. Gärdenfors uses feature based cognition as his basis. In feature based cognition, an object is not recognized because it is similar in some degree to a prototype of a certain class, but because it has a certain amount of features in common with other objects that do belong to one class [Sternberg, 2003].

[2]Although Kamp and Partee use a measure function for both the degree of prototypicality and the degree of membership, the main ideas behind their approach are of prototype based cognition rather than feature based cognition.

Therefore, Gärdenfors constructs in his article conceptual spaces out of different quality dimensions. These represent the different features, for example colour, weight, shape, etc. In these spaces, it is possible to represent the qualities of different objects in different domains. Gärdenfors considers the "set of integral dimensions that are separable from all other dimensions" [Gärdenfors, 1998] as a domain, where the integral dimensions are the ones that cannot be separated. This means that if it is the case that if an object receives a value in dimension A, it must also receive a value in dimension B, then A and B are integral dimensions.

An example of two integral dimensions are the hue and the saturation of colour. If we give an object a value for the hue of the colour, then we also have to give it a value for the saturation. Therefore these two dimensions are not separable and thus integral dimensions. Every dimension that is not integral with another dimension is said to be separable.

An example of a domain is the colour domain. When we use all the dimensions that are integral with the respect of colour information, we get a 'colour spindle'.

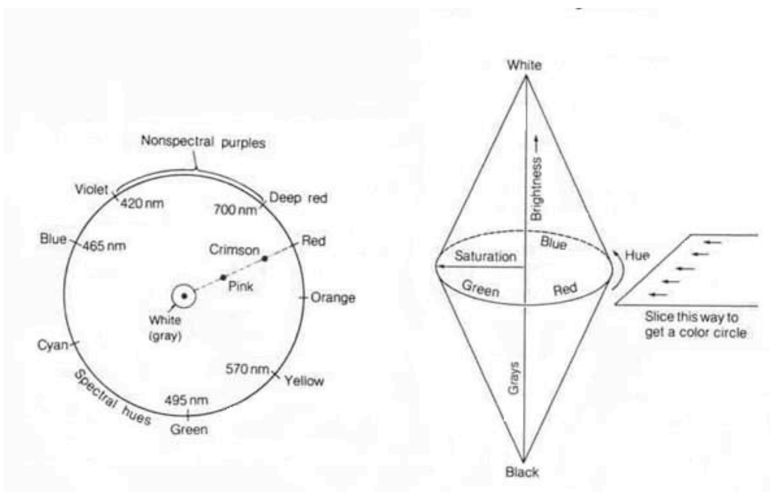


Figure 5.2: The 3D colour spindle.

In this picture, the well known colour circle makes the middle of the cone. The saturation goes from the axis of the cone (unsaturated) to the outer border (fully saturated). The hue goes around the border (as it does with the colour circle) and the brightness goes from the top down, along the axis.

There is some psychological evidence that human beings do construct this kind of domains in their mind. It is, however, unclear which dimensions are innate and which are learned during life.

With these domains, it is possible to describe the different properties of an object. So, for example, the property ‘red’ of a ‘red apple’, is a convex region within de colour domain. Properties, like ‘red’, ‘small’ or ‘fat’, are special kinds of concepts, because these properties only refer to a single domain, where normal concepts can refer to multiple domains[3].

Although concepts can refer to multiple domains and therefore are constructed out of different properties, they are not just a mere collection of these properties. Instead, concepts are constructed of properties and some correlation between those properties. Or, as Gärdenfors puts it:

> “A natural concept is represented as a set of convex regions in a number of domains together with a prominence assignment to the domains and information about how the regions in different domains are correlated.” [Gärdenfors, 1998]

5.2.1 Gärdenfors’s Concept Combination

The result of this geometrical model is the following: When we want to combine two concepts, such as ‘red’ and ‘apple’, we will replace the colour property of apple, with the colour property of red.
In this case, it is fairly simple, because ‘red’ is a simple concept (it refers only to one property) and because the noun takes on the only property of the adjective. Note that this notion of concept combination is very different from the Boolean algebra used in concept combination in the articles of Osherson & Smith and Kamp & Partee.

Unfortunately, it is not always this simple. If we look at the case of the ‘stone table’, several problems arise. First of all, ‘stone’ is not a simple concept. Therefore, the noun cannot simply replace one of his own properties with the one that the adjective refers to[4]. It gets even more difficult when we look at the case of ‘stone lion’, where the adjective receives one of the

[3]These are generally the more complex concepts like ‘cat’, ‘female’, etc etc.

[4]for another good example, see the ‘foolish bird’ example of Armstrong, Gleitman and Gleitman, [Armstrong et al., 1983], p.272

properties of the noun, instead of the other way around.

As we said before, it is not really clear when the noun takes on the property (or properties) of the adjectives and when it is the other way around. Kamp and Partee already described this as context dependent ([Kamp and Partee, 1995], p. 142 and 143).

> "relative adjectives like tall, heavy and old are context-dependent as well as vague, with the most relevant aspect of context a comparison class which is often, but not exclusively, provided by the noun of the adjective-noun construction."
> ...
> "It is both difficult and important to try to sort out the effects of context dependence on the interpretation of different sorts of adjectives and nouns, both alone and in combination."

Another problem is the problem of relativity. If we describe a 'red book' and a 'red apple', we use the same word ('red'), but refer to different colours. Gärdenfors solves this by introducing contrast classes. These contrast classes are subclasses of a domain. Imagine the colour spindle used to exemplify the colour domain. In this domain, there will be a (convex) sub domain containing the colours of an apple. This way, if we make a reference to a 'red apple', we will refer to all the red colours in the apple colour domain. Gärdenfors says this as follows:

> The combination XY of two concepts X and Y is determined by letting the regions for the domains of X, confined to the contrast class defined by Y, replace the corresponding regions for Y.

This seems like a rather good solution for the 'red apple' or 'red nose' case. Just intersect the colour regions of the two concepts and take the intersection as a new colour region for the new formed combinatorial concept. It is a rather intuitive notion as well. For example, when we talk about a photo, we will probably see a full coloured picture in our mind. When we have an information update, that tells us that it is actually a black and white photo, we will not change the image that we thought was displayed on it, but we will simply imagine that photo, but now without any colour (except for black, white and different shades of grey).
It is, however, still a rather sketchy solution, that Gärdenfors does not specify.

5.3 Conclusion

In his articles and book, Gärdenfors presents some really nice ideas on representing concepts in a geometrical model.
However, he does not specify how this model should be implemented and how it should be used. Because he uses convex concept spaces to represent his concepts, there is no straightforward way to represent his model in a more formal way, in order to see whether his concept combination actually works and whether it can handle the conjunction fallacy.

In the next chapter, I will discuss another way of representing concepts in a geometrical space: the quantum approach.

Chapter 6

Concepts in geometrical models – the quantum approach

Gärdenfors's model looks fairly promising, but unfortunately is there no straightforward natural way of implementing it. Still, there is a way to represent concepts in a geometrical model (in this case: a Hilbert space) that can be formalized: the quantum approach.
Here, concepts will be represented as vectors in a Hilbert space and we will use a quantum based probability theory instead of the classical probability theory.
In this chapter we will show that using quantum probability theory and geometrical models this way, this quantum approach can handle the conjunction fallacy.

6.1 How did it work in classical probability theory?

Before we can go into the mathematical implementation of the quantum approach, we will take a closer look at the classical Boolean algebra. In this section we will show what classical probability theory can do and what its flaws are.

Consider two observable concepts ('Striped' (S) and 'Apple' (A)), each of which can have two values (yes or no, e.g. something is or is not an apple). The process of observing the concepts is showed in figure 5.3.

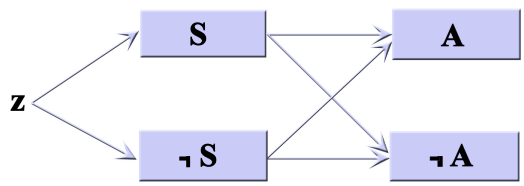


Figure 6.1: Process of observing the two concepts [Busemeyer et al., 2006]

First, we are in an initial state z. When we observe the object, we will find if it is striped or not (S or $\neg S$). Next, we will see if it is an apple or not (A or $\neg A$). Now, all the possible outcomes are SA, $S\neg A$, $\neg SA$, $\neg S\neg A$. These outcomes are mutually exclusive (e.g. $SA \wedge S\neg A = \emptyset$) and exhaustive (all the possible outcomes form together the whole set U). New concepts (like 'striped apple') can now be formed by negation and combination of the two original concepts. For combining the two concepts into one, the normal Bernoulli algebra is applied and the Kolmogorov rules hold for the different probabilities:

Bernoulli algebra:

- Commutative: $x \vee y = y \vee x$
- Associative $x \vee (y \vee z) = (x \vee y) \vee z$
- Complementation: $x \vee (\neg y \vee y) = x$
- Absorption: $x \vee (x \wedge y) = x$
- Distributive: $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$

Kolmogorov rules:

- $0 \leq P(x) \leq 1$, $P(\emptyset) = 0$, $P(U) = 1$
- If $x \wedge y = \emptyset$ then $P(x \vee y) = P(x) + P(y)$

Now we also know that:

- $P(A) = P(A \wedge (S \vee \neg S)) =$
  $P((A \wedge S) \vee (A \wedge \neg S)) =$
  $P(A \wedge S) + P(A \wedge \neg S)$

- $P(A|S) = \frac{P(A \wedge S)}{P(S)}$
- therefore: $P(A \wedge S) = P(S) \cdot P(A|S)$

We know that because $P(A|S) \leq 1$, the value for $P(A \wedge S)$ will be smaller or equal to P(S)[1]. This is the conjunction fallacy. But there is more that goes wrong when using classical set theory like this, and later I will show that these faults are connected to the conjunction fallacy.

When one sees an object 'apple', one can conclude A or $\neg A$. So there is a chance that one concludes A when one sees an apple ($P(A|apple)$). But there is no rule that requires that[2] $P(A|striped) = P(A|striped; apple)$.

However, if these two terms are not equal, they have a huge psychological effect. The difference between $P(A|striped)$ and $P(A|apple; striped)$ is called a forward interference effect, also called priming. For example, when one sees a picture of a bike for a very short period (a tenth of a second), the subjects will faster recognize the word 'bike' when that is presented afterwards, than the subjects who did not see the picture. The difference between $P(A|striped)$ and $P(A|striped; apple)$ is called the backward interference effect, also called masking. When subjects get to do a test question, they perform worse if they are shown a 'noise sample' afterwards than if they are not.

Concluding, we have tried for several decades to explain (and predict) human behavior by applying classical logic and classical probability theory to their actions. It turns out that these methods have a few flaws. One of these flaws is the conjunction fallacy.
The next section will show how using quantum probability theory instead of classical probability theory can help to solve the conjunction fallacy.

6.2 Quantum probability theory

In this section I will give a short introduction to quantum probability theory. This is a generalization of classical probability theory and based on the geometrical models that originate in physics.

[1] A and S are independent, this still holds. Remember that the formula in that case is $P(A \wedge S) = P(S) \cdot P(A)$

[2] When we are observing that something is an apple first, and later that it is striped, we write $A; S$. $P(A; S) = P(A) \cdot P(S|A)$

Quantum probability theory applies to two different groups: compatible measures and incompatible measures. Compatible measures are cases where two dimensions can be measured, accessed or experienced simultaneously [Busemeyer and Wang]. An example of compatible measures are the integral dimensions mentioned by Gärdenfors. For example, the hue and saturation of a colour can be experienced and measured at the same time and are therefore compatible measures.
Because Quantum probability theory works the same for compatible measures as classical probability theory, we will not go into this subject any further[3]. We are more interested in the incompatible measures, because that is where the problems arise when we are trying to combine two concepts.

Incompatible measures are measurements that cannot be done simultaneously. An example is 'stone lion'. Here, we cannot process the measurement of the material (stone) and the shape (lion-shaped) at the same time. These measurements will be done, one after another.

Now let us look a bit deeper into the technical details of quantum probability theory.
Every agent, which is about to experience something is in a certain state already. In quantum probability theory, this is represented by a state vector z, that is an element of a Hilbert space $\mathcal{H}$. We make sure that the length of the state vector is 1: $< z|z >= 1$. Note that we are using the bra-ket notation introduced by Dirac [Dirac, 1982].
Every state vector is based on several events. Every event is represented by a vector and the state vector is a superposition of these different events (it represents the experiences of the agent). In the next section, we will go deeper into superposition.

In the Hilbert space, every observable is represented by orthogonal vectors. Let us assume that we only observe the concept 'apple'. Now, we have two vectors x (representing A) and y (representing $\neg A$), that are orthogonal to each other (So $< x|y >= 0$). In other words, they span a plane in $\mathcal{H}$. This can be seen in figure 5.4.

Further, we make sure that $< x|x >=< y|y >= 1$. For every event L_x there is a corresponding projection operator Π_x that makes a projection of the vectors in $\mathcal{H}$ onto L_x. For example, it is possible to make a projection from the state vector to L_x. This represents the chance that z collapses into L_x. This is shown in figure 5.4.

[3]For more information on this subject, check the article of Busemeyer and Wang [Busemeyer and Wang]

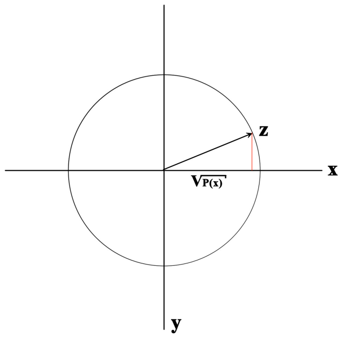


Figure 6.2: Orthogonal vectors y and x and state vector z.

The projection operator of the whole Hilbert space ($\Pi_{\mathcal{H}}$) is equal to the identity matrix I. This is the case because $\Pi_x = |x><x|$ and $\Pi_y = |y><y|$.
$\Pi_{\mathcal{H}} = \Pi_x + \Pi_y = |x><x| + |y><y| = I$.
The chance that z collapses into L_x is the same as $P(x)$. The square of the length of the mapping of the state vector onto L_x is $P(x)$. So:

$$\begin{aligned} P(x) &= (\Pi_x|z>)^\dagger(\Pi_x|z>) \\ &= |\Pi_x|z>|^2 \end{aligned} \tag{6.1}$$

The squares of $<x|z>$ and $<y|z>$ are the probability amplitudes such that $|z>$ collapses into $|x>$ or $|y>$. Therefore, $<x|z>$ and $<y|z>$ are the coördinates that represent the initial state. They can be any complex number, such that $|<x|z>|^2 + |<y|z>|^2 = 1$, because $<z|z>= 1$.

Now let us assume that we do not only measure just one observable, but two. Remember that x represented A and y represented $\neg A$. Now u represents S and v represents $\neg S$. Now we get a different picture, because we can choose which set (xy or uv) we choose to set as our basis. These four axes share the same Hilbert space, but are rotated. This way, they span the same plane, but are still orthogonal. This is shown in figure 5.5.

Now we get that:

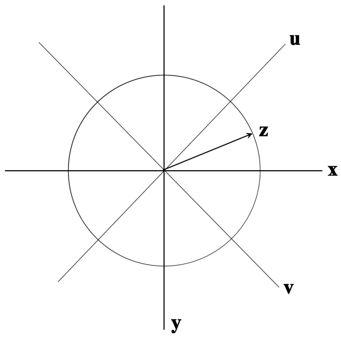


Figure 6.3: Orthogonal vectors x, y, u and v and state vector z. [Busemeyer and Wang]

$$\begin{aligned}|\Pi_{xu}|z>|^2 &= |<xu|z>|^2 \\ |\Pi_x|z>|^2 &= |<xu|z>|^2 + |<xv|z>|^2\end{aligned} \tag{6.2}$$

In classical probability theory, it does not matter whether we process the concept 'apple' first, or the concept 'striped'. But as Frank Veltman showed in his Nicholas example [Veltman, 1998], in human reasoning, it does matter[4].
This is one of the things that quantum probability theory can handle. Because:

$$\begin{aligned}\Pi_x \cdot \Pi_u &= |x><x| \cdot |u><u| =<x|u>|x><u| \\ \Pi_u \cdot \Pi_x &= |u><u| \cdot |x><x| =<u|x>|u><x| \\ <x|u>|x><u| &\neq <u|x>|u><x|\end{aligned} \tag{6.3}$$

Another way to show this is by figure 5.6. Here, one can see the difference between choosing xy or uv as a basis. If one observes the concept 'apple' first, one (unconsciously) chooses xy as a basis. As long as $|<x|z>|^2$ is equal to $|<u|z>|^2$, there will be no difference between choosing the xy or uv as a basis. But as you can see in the picture, there clearly is a difference in this case. Therefore, the order in which things are observed is still important.

[4]Remember the notiation $(A; S)$, which meant that one observes that the object is an apple first, and then observes that the object is striped

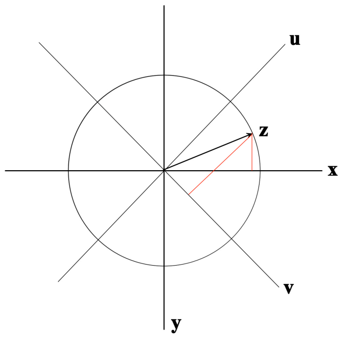


Figure 6.4: The difference between $| < x|z > |^2$ and $| < u|z > |^2$ [Busemeyer and Wang]

If we want to know the chance that the vector z collapses into the vector u ($| < u|z > |^2$), we need to know first the chance of the vector of u given x and given y. So the resulting formula will be:

$$\begin{aligned} | < u|z > |^2 = \quad & | < u|x > |^2 \cdot | < x|z > |^2 + | < u|y > |^2 \cdot | < y|z > |^2 + \\ & 2| < u|x >< x|z > | \cdot | < u|y >< y|z > | \cdot cos(\theta) \end{aligned} \tag{6.4}$$

Note that this formula is different from the formula that is obtained from classical probability theory[5]. The difference lies in the last term. This term is called the 'inference term'. It describes the angle between $< u|x >< x|z >$ and $< u|y >< y|z >$. If this angle is 0, there will be no interference.

This interference term is the cause of the conjunction fallacy. As Riccardo Franco clearly showed in his article 'The conjunction fallacy and interference effects' [Franco, 2007], if $cos(\theta)$ is negative, $P(u)$ can be lower than $P(u|x)$. Which equals the chance of first observing x and then observing u ($P(x; u)$). This is the basis of the conjunction fallacy

6.2.1 Conclusion

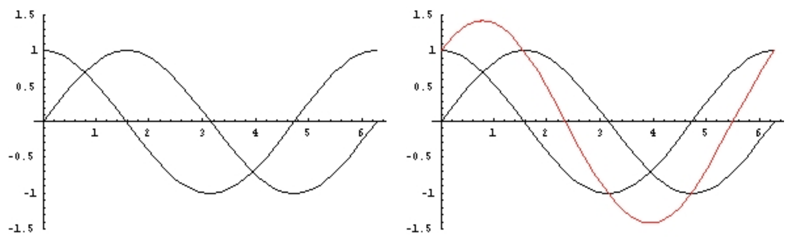
This section showed that quantum probability theory differs from classical probability theory in the way that incompatible measures are handled. We have seen that the order in which concepts are observed changes the way we interpret them.

[5] $P(u) = P(u|A) + P(u|\neg A)$

In the next subsection, we will take a closer look at superposition, to get a clearer view on the idea behind using quantum probability theory in concept combination.
In the last section of this chapter, we will show that quantum probability theory really is the answer to our questions.

6.3 A bit more on superposition

At this point, we are going to take a little time out to take a closer look at superpositioning.
Superpositioning is a term that has its origin in quantum mechanics. It happens when two waves coincide. Now the amplitudes of the waves are added[6] to each other, forming one new wave. Sometimes, two waves can even cancel each other out (e.g. when their amplitudes add up to zero).



(a) Two waves (b) and their superposition wave.

Figure 6.5: Two waves and their superposition wave.

As you can see in the picture, the waves cancel each other out at x $\approx$ 2.3. But at x $\approx$ 0.8, the superposition wave is at his top, where the two basic waves are not.
This is precisely what we are going to use when we are trying to solve the conjunction fallacy. Now the resulting c^p value can be higher than both the basic c^p values of both the conjuncts.

In human cognition, superposition often occurs. This is called the interference effect. For example, when a subject stares for a certain amount of time at a white paper with a black dot on it and afterwards he looks at a completely blank paper, he will still see the dot, although it is not there. This is called the afterimage. The dot that one sees is the opposite colour

[6]To be more precise: a linear operation is performed on the original waves, to create a new wave.

of the original dot.
If one stares for a really long time at a black dot, and than looks at a piece of paper that is half red and half white, the dot (that was on the edge of both colours) will now have a different colour on each half. This is caused by superposition of the previous image with the present image (a completely blank paper at one half of the dot and a red paper at the other half).

Also the state a human being is in, can be seen as a superposition of all his experiences. This is one way to create 'common knowledge'. Now, the experiences will not simply be added to each other, because not every experience is as important as another experience. Therefore, there will be some (weighted) linear operation that forms this superposition wave.

6.4 Concept combination with quantum probability theory

In the last two sections, we took a closer look at the quantum probability theory. The question now remains: does it actually work? That is what we are going to find out in this section.

Let us shortly recapitulate what the problems that needed to be solved were:

1. The universally true and universally false sentences
2. The conjunction fallacy

In the next subsections, I will discuss both of the problems.

6.5 quantum probability theory and the universally true and false sentences

Let us first consider the universally true sentence. This is a sentence like: Milo is a cat or not a cat. Now the question is: are the chances that Milo is a cat or is not a cat one? Or, differently phrased, is it true that $\Pi_{x \vee y} = 1$ (where x stands for 'Milo is a cat' and y stands fore 'Milo is not a cat')? Note that we are now calculating the chance that the vector z collapses in either y or x. This is (in this case) the same as calculating the chance to x

or y $(P(x \vee y))$
The answer to this question must obviously be 'yes'. Because x and y span the whole two-dimensional Hilbert space (H), $x \vee y = \mathcal{H}$. And $P(\mathcal{H}) = 1$.

The universally false sentence is equally trivial. Because the vectors x and y are orthogonal, the chance that $x \wedge y$ is true, must obviously be 0.

These problems are solved. Now let us see if the conjunction fallacy can be solved as easily as the universally true and false sentences.

6.6 quantum probability theory and the Conjunction Fallacy

Recall definition (5.4). This formula contains the answer to the question whether or not it is possible to explain the conjunction fallacy.

The last part of this formula contained the interference term. This interference term is the cause of the conjunction fallacy. As Riccardo Franco clearly showed in his article 'The conjunction fallacy and interference effects' [Franco, 2007], if $cos(\theta)$ is negative, $P(u)$ can be lower than $P(u|x)$. Which equals the chance of first observing x and then observing u $(P(x; u))$. This is the basis of the conjunction fallacy.

Let's explain it a bit more thoroughly. Recall that our state $|z>$ equals $|x><x|z> + |y><y|z> = |u><u|z> + |v><v|z>$
We know that $P(u) = P(u \wedge (x \vee y))$. Therefore [Busemeyer and Wang]:

$$\begin{aligned} P(u) &= P(u \wedge (x \vee y)) \qquad (6.5) \\ &= |<u|x><x|z> + <u|y><y|z>|^2 \\ &= |<u|x><x|z>|^2 + |<u|y><y|z>|^2 + 2|<u|x><x|z>| \\ &\quad \cdot |<u|y><y|z>| \cdot Cos(\theta) \qquad (6.6) \end{aligned}$$

So: if $Cos(\theta) < 0$ it is possible that $P(u) < P(u|x)$ (6.7)
which causes the conjunction fallacy.

6.7 Conclusion

In this section we explained how geometrical models can be implemented with quantum probability theory. In this section it was made very clear that using quantum probability theory instead of classical probability theory solved the problems of the universally true and false sentences and the conjunction fallacy in a fairly easy and natural way.

Part III

Are we on the right track?

Chapter 7

Conclusion and Future work

7.1 Conclusion

In this thesis I have presented several attempts to solve the conjunction fallacy and they were all more or less successful.
We have seen that fuzzy set theory [Osherson and Smith, 1981] and supervaluation theory [Kamp and Partee, 1995] managed to make a big step in the right direction, but could not solve the problem in all cases[1].
The methods by Veltman [Veltman, 1998] and by Gärdenfors e.g.[Gärdenfors, 1998] presented us a whole new way of thinking, that seemed to provide the solution to our problems. Unfortunately, they were too abstract to implement directly.
Finally we discussed the articles of Busemeyer, Wang and Townsend [Busemeyer et al., 2006] and Franco [Franco, 2007], where quantum probability theory was presented as the method that would solve all our problems concerning concept combination.
We have seen that, at least in the case of the conjunction fallacy, the quantum approach did actually work! This was a result that was surprising, but also very promising, because it opens up a whole new approach to the research of human reasoning.

We have also seen that not all the researchers interpret the conjunction fallacy as the same problem. Some treat the conjunction fallacy as a problem of classical probability theory and others as a resemblance problem.
The main question now remains: does the use of quantum probability theory even handle this problem?

[1]The different cases that we considered were: the universally true sentences, the universally false sentences and the 'striped apple' case.

7.1.1 Is the right problem solved?

Did the implementation of quantum probability theory really solve the right solution? In other words: what version of c^p is solved? What about the other one?

Let us recall the two versions of c^p. The first version really expressed the chance that something was the object that was mentioned. For example $c^p_{Apple}(a)$ will express in this case the chance that object a is an apple.
The second version of c^p was a measure of resemblance. Here, $c^p_{apple}(a)$ expresses the measure in which a resembles an apple.

Frank Veltman's default logic system was based on the resemblance explanation of c^p. In his default rules, he basically formalized the rules of resemblance, but had no probabilistic measure.

Also Gärdenfors's geometrical model was primarily based on the second version of c^p. He created a model where a certain feature –for example colour– replaced the feature of the original object. For example, when we talk about a 'red apple', we replace, according to Gärdenfors, the colour domain of the apple, with the intersection of the colour domain of the apple, and the colour domain of the term 'red'. Resulting in a 'red apple colour domain'.

These two approaches seemed to be the solution to the problems encountered by Osherson & Smith and Kamp & Partee. But does the implementation with quantum probability theory still only solve the problems for the second version of c^p?
We claim that it does not. The use of quantum probability theory solves the problems of both versions of c^p. When using quantum probability theory instead of classical probability theory, one gets the same results as subjects get when using the second version of c^p. In other words: it is an adequate model of human behavior.
It uses the chances that an object is for example an apple. And it is a really straightforward implementation of a probability theory. Therefore, this implementation also works for the first version of c^p, where $c^p_{Apple}(a)$ is the chance that a is an apple and not the measure of how similar a is to an apple.

So, concluding, when we use quantum probability theory in a geometrical model, the problem of the conjunction fallacy can be solved in a principled way, for both versions of c^p.

7.2 Future work

Now that the quantum approach has proven to be a fruitful one, it opens many doors to solutions to other problems that are concerned with bounded rationality. Here, I will give a short, and by no means exhaustive, list of other problems that can be solved by using the quantum approach:

- **The disjunction fallacy** The disjunction fallacy states that the probability of AB is smaller than both the probabilities of A and B. Intuitively, this can be explained with superposition, just as the conjunction fallacy.
- **The conditional probability fallacy** The conditional probability fallacy states that $P(A|B)$ is approximately equal to $P(B|A)$. This is an assumption often made, but in many cases not true.
- **The framing effects** Framing effects occur when a situation is placed within a scenario. When one needs to answer a question, the context of that question has an impact on the answer. It seems plausible that this can be explained in the same way as the conjunction- and the disjunction fallacy.
- **The ordering effects** It does matter in what order you say something to another person (think of Veltman's Nicholas example). This is called the ordering effect. Because quantum probability theory takes the order in which things are observed into account –in contrary to classical probability theory– I would presume that the quantum approach can be helpful here as well.
- **The Elssberg paradox** If you have a bucket with 90 balls, 30 of which are white and sixty of which are black or red. If one has to choose between (a) you will win if you pick a white ball and (b) you will win if you pick a black ball, he will always pick option (a) iff he believes that the chance of picking a black ball is smaller than picking a white ball. But if this is the case, he should also prefer (c) you will win if you pick a white or a red ball over (d) you will win if you pick a black or red ball. But in reality, many people prefer (a) over (b) and (d) over (c). This is the Elssberg paradox.
- **The Allais paradox** This paradox occurs –just as the previous paradox– when people do not act according to utility theory. If one has to choose between (a) 100% chance of winning 1 million dollars and (b) 89% of winning 1 million dollars, 1% of winning nothing and 10% of winning 5 million dollars, most people would prefer option (a). But if he has to choose between (c) 89% of winning nothing and 11% of winning 1

million and (d) 90% of winning nothing and 10% of winning 5 million, most people prefer option (d). That the same person would choose both (a) and (d) is inconsistent with expected utility theory.

Acknowledgements

I would like to thank my supervisor Reinhard Blutner, for his inspiration, advice and time that he gave me when I was writing my thesis.
Second, I would like to thank Frank Veltman for a very helpful discussion halfway the process of writing my thesis.
And last, but not least, I would like to thank Paul, my mum and dad and my brothers for encouraging me to write this thesis and *finally* finishing it.

Bibliography

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

J.R. Busemeyer and Z. Wang. Introduction to quantum information processing for cognitive scientists: Quantum probability. *to appear*.

J.R. Busemeyer, Z. Wang, and J.T. Townsend. Quantum dynamics of human decision-making. *Journal of Mathematical Psychology*, 50:220–241, 2006.

P.A.M. Dirac. *The principles of quantum mechanics*. Oxford University Press, fourth edition edition, 1982.

K. Fine. Vagueness, truth and logic. *Synthese*, (30):265–300, 1975.

B.C. van Fraassen. Singular terms, truth-value gaps and free logic. *The journal of Philosophy*, 63(17):481–495, 1966.

R. Franco. The conjunction fallacy and interference effects. *arXiv*, August 2007. URL `http://xxx.lanl.gov/pdf/0708.3948`.

P. Gärdenfors. Concept combination: a geometrical model. 1998.

P. Gärdenfors. Conceptual spaces as a basis for cognitive semantics. *Appeared in the proceedings of Second International Colloquium on Cognitive Science, Universidad del Pais Vasco*, 1993.

P. Gärdenfors. Meanings as conceptual structures. 1995.

P. Gärdenfors. *Conceptual Spaces: the geometry of thought*. MIT Press, Cambridge, 2000.

D. Kahneman. Maps of bounded rationality: A perspective on intuitive judgment and choice. *Nobel Prize Lecture*, 2002.

H. Kamp and B. Partee. Prototype theory and compositionality. *Cognition*, (57):129–191, 1995.

W. Labov. *New ways of analyzing variation in english*, chapter Part 1, pages 29–62. University Press, 1973.

J. van Ligtenberg. Evaluating theories of concept combination. *Bachelor Thesis*, november 2007.

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

Plato. The republic. (7):5.14a.

R. Reiter. A logic for default reasoning. *Artificial Intelligence*, pages 81–132, 1980.

E. Rosch. *Concepts: core readings*, chapter 8, pages 189–205. MIT Press, 1999.

R.J. Sternberg. *Cognitive Psychology*. Wadsworth, third edition edition, 2003.

K. Tentori, N. Bonini, and D. Osherson. The conjunction fallacy: a misunderstanding about conjunction? *Cognitive Science*, 28:467–477, 2004.

the Stanford Encyclopedia of Philosophy. Concepts. 2007. URL `http://plato.stanford.edu/entries/concepts/#ProThe`.

A. Tversky and D. Kahneman. Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, 90: 293–315, 1983.

F.J.M.M Veltman. Een zogenaamde denkfout. *Algemeen Nederlands Tijdschrift voor Wijsbegeerte*, 90:11–26, 1998.

L.A. Zadeh. Fuzzy sets. *Information and control*, (8):338–353, 1965.