University of Tartu
Institute of Philosophy and Semiotics

A Classical Degree-Theoretic Treatment of the Sorites Paradox

Master's Thesis in Philosophy

Ching Hei, Yau

Supervisors:
Tuomo Juhani Yli-Vakkuri,
Bryan Reed Stewart Frances

2019

For Bryan and Juhani

Table of Contents

Chapter 1. Background and Motivations[1]

Since 1970s, degree-of-truth theory has been proposed as a solution to the Sorites paradox.[2] However, one perennial attack to degree-of-truth theory is that its logic - fuzzy logic - is *non-classical*.[3] Inspired by Gödel (1933), I attempt to better degree-of-truth theory by *classicalizing* it. That is, I attempt to give an interpretation of fuzzy logic within classical logic enriched by degree operators {○, ◔, ◑, ◕, ●} - "it is of no/low/moderate/high/full degree that…". Intuitively, degree-of-truth is classicalized as classical bivalent truth-value and a largely independent notion of degrees. A formal semantics of this enriched classical logic is presented, from which two semantic consequences are derived. The two semantic consequences are applied to analyse the *(in)validity* of the Sorites argument. There are two results: 1. the validity of the standard Sorites argument is reasserted, 2. a new argument for the *in*validity of the *degreed* version of the Sorites argument is presented.

In this chapter, I will provide the historical background and philosophical motivations of this thesis. Chapter 1.1 presents the standard Sorites paradox. Chapter 1.2 presents degree-of-truth theory (hereafter "degree theory") and its solution to the Sorites paradox. Chapter 1.3 presents Gödel's classicalization of intuitionistic logic. Chapter 1.4 presents our methodology. Chapter 1.5 outlines the structure of this thesis.

1.1 The Sorites Paradox[4]

Below is a standard *Sorites argument*:

---

[1] Professor Juhani Yli-Vakkuri and Professor Bryan Frances have contributed to many parts of this thesis. They have my sincerest gratitude for their world-class supervision, unending support, and encouragement.

[2] Most notably, by Goguen (1969) "The Logic of Inexact Concept", Lakoff (1973) "Hedges: a Study in Meaning Criteria and the Logic of Fuzzy Concepts", Machina (1976) "Truth, Belief, and Vagueness", King (1979) "Bivalence and the Sorites Paradox", Forbes (1983) "Thisness and Vagueness", Edgington (1992) "Validity, Uncertainty and Vagueness", Edgington (1997) "Vagueness by Degrees", Hyde (2008) *Vagueness, Logic and Ontology*, and Smith (2008) *Vagueness and Degrees of Truth*.
There was an explosion of interest in vagueness around the 1970s. According to Edgington 1992 p.199, degree theory *was* one of "the two best-known solutions" to the Sorites paradox (, the other being supervaluationism). Degree theory is still very much alive. In edited Dietz and Moruzzi (2009) *Cuts and Clouds: Vagueness, Its Nature, and Its Logic*, published by Oxford University Press, 5 of its 31 papers are devoted to "Many-Valued Logics".

[3] See e.g. Smith 2008 chapter 5.2, Williamson 1994 p. 118, and McGee and McLaughlin 1995, p. 237.

[4] The ancient Greek adjective "*sorites*" comes from the noun "*soros*", which means heap. "*Sorites*" literally means *heaper*, so the Sorites paradox is "the heaper paradox" (Williamson 1994 p. 9).

Premiss 1:	10,000 grains of sand make a heap.
Premiss 2:	If 10,000 grains of sand make a heap, then 9,999 grains of sand make a heap.
Premiss 3:	If 9,999 grains of sand make a heap, then 9,998 grains of sand make a heap.
…	
Premiss 10,000:	If 2 grains of sand make a heap, then 1 grain of sand makes a heap.
Premiss 10,001:	If 1 grain of sand makes a heap, then 0 grain of sand makes a heap.
Conclusion:	0 grain of sand makes a heap.
Rule of inference:	10,000 instances of *modus ponens*.[5]

The Sorites argument is a *paradox* in the sense that it is an apparently valid argument, with apparently true premisses, but an apparently false conclusion. Apparently, there are three possible solutions, jointly exhaustive:

1. A premiss is false.
2. The argument is invalid.
3. The conclusion is true.

1.2 Degree-of-Truth Theory

The law of excluded middle (LEM) is at the core of classical logic. LEM provides that every proposition is either true or false.[6] Degree theorists reject some instances of LEM, and thereby reject classical logic. Instead of LEM, degree theorists hold that *propositions are true to different degrees*. For example, degree theorists would say that Plato and Whitehead are both influential philosophers, but the proposition that *Plato is an influential philosopher* is true to perhaps the highest degree, while that *Whitehead is an influential philosopher* is true to a lesser degree.

[5] *Modus ponens* is the inference from $\{\varphi, (\varphi \rightarrow \psi)\}$ to ψ.

[6] This is the traditional (i.e. Aristotle's and Russell's) definition of LEM. LEM now sometimes refers to the schema "$(\varphi \vee \sim\varphi)$" in the object-language under study. The *principle of bivalence* is the meta-linguistic principle which says that any well-formed formula (wff) "φ" in the object-language is either true or false (Williamson 1994 p. 145). In this thesis, I will simply adopt the traditional definition.

Standard degree-of-truth theorists[7] use their non-classical logic to argue that the Sorites argument is *invalid*:

1. Firstly, they define validity as the *preservation of degree-of-truths*, from the premisses, to the conclusion, in all models. Accordingly, if and only if a conclusion could be less true than the least true premiss, then the argument is invalid.[8]
2. Secondly, they assign each Sorites premiss with a very high degree-of-truth, and assign the Sorites conclusion with a very low degree-of-truth. So *the Sorites conclusion is less true than the least true Sorites premiss*.
3. By 2, they conclude that the Sorites argument is *invalid*, according to their definition of validity given in 1.

However, this proposed solution to the Sorites paradox comes at the cost of:

1. *rejecting the law of excluded middle*, as degree theorists hold that propositions are not simply either true or false, but are true (or false) to different degrees;[9] and

---

[7] According to Smith 2008, p.220.

[8] This was already done in Machina (1976 p. 63): "[when] its conclusion must always be at least as true as the falsest premise, the argument will be fully valid."

[9] To be exact, the fact is that degree theorists issue less than full truth to some instances of the law of excluded middle ($\varphi \vee \sim\varphi$), and to some instances of the law of non-contradiction $\sim(\varphi \,\&\, \sim\varphi)$.
According to standard degree theory, the degree-of-truth °T:

$$°T(\sim\varphi) = 1 - °T(\varphi)$$

$$°T(\varphi \vee \psi) = \max.\{ °T(\varphi), °T(\psi) \}$$

$$°T(\varphi \,\&\, \psi) = \min.\{ °T(\varphi), °T(\psi) \}$$

For example, if $°T(p) = 0.7$ and $°T(q) = 0.3$, then

$$°T(p \vee q) = \max.\{ °T(p), °T(q) \} = 0.7$$

$$°T(p \,\&\, q) = \min.\{ °T(p), °T(q) \} = 0.3$$

Then, when $°T(p) = 0.5$,

$$°T(\sim p) = 0.5$$

$$°T(p \vee \sim p) = 0.5$$

$$°T(p \,\&\, \sim p) = 0.5$$

$$°T\sim(p \,\&\, \sim p) = 0.5$$

That is to say, an instance of the law of excluded-middle is only half true, a contradiction is half true, and an instance of the law of non-contradiction is only half true.

2. *rejecting the validity of modus ponens* as a rule of inference, as the *only* type of rule of inference used in the Sorites paradox is *modus ponens*.

Since 1970s, degree-of-truth theory has been proposed as a solution to the Sorites paradox. However, since Kit Fine (1975)'s supervaluationism, and Timothy Williamson (1994)'s epistemicism, degree theory has faced one perennial attack. The perennial attack to degree theory is this: without clear outweighing theoretic advantage, *degree theory does not fully preserve classical logic*. For many analytic philosophers, the revision of classical logic is very costly, and should not be done without knowing clearly 1. what would be gained, and 2. that the gain outweighs the cost.

Nevertheless, degree theory is well-motivated: There are some *facts*, which made, and still make, degree theory appealing; for example, the fact that *red things are red to different degrees*. This is a fact that even diehard classical logicians admit, and need to accommodate it in their classical logic.

While the fact that fuzzy logic rejects LEM is undeniable, it is not necessary to reject LEM in order to talk about degree and truth. If the worst thing about degree theory is that its logic is non-classical, then we can better degree theory by making its logic classical. As Gödel shows us, this can largely be done.

1.3 Gödel (1933)'s Classicalization of Intuitionistic Logic

Intuitionistic logic also rejects LEM. Intuitionists think that what we should say about mathematical statements, such as "$0 - 1 = -1$", is *not* whether they are true in the sense of word-world correspondence (what are the worldly correspondents of "0" and "−1"?), but whether that statement is *provable*. According to intuitionists, any admissible mathematical statement p has one of the following three "proof-conditions", jointly exhaustive:

1. p: It is provable that p.
2. $\neg p$: It is provable that not-p.

3. None of the above.[10]

In short, intuitionistic logic has a non-classical notion of negation. That is, the truth-value of "¬φ" is *not* truth-functionally determined by the truth-value of "φ". For example, when the truth-value of "φ" is false, the truth-value of "¬φ" is not thereby determined, such that the truth-value of "¬φ" can still be true, and can still be false. In intuitionistic logic, if *p* is the twin prime conjecture, then "*p*" is false and "¬*p*" is false, so it is not the case that "($p \vee \neg p$)". But to say it is not the case that "($p \vee \neg p$)" is to reject an instance of LEM. So intuitionists reject some instances of LEM (Priest 2008 p. 104). This non-truth-functional notion of negation, and this rejection of some instances of LEM, are what we primarily mean, when we say that intuitionistic logic is "non-classical".

In "An interpretation of the intuitionistic propositional calculus" (1933), *Gödel gave an interpretation of intuitionistic logic within classical logic enriched by an operator B "It is provable that …"*.[11] Gödel proposed a translation scheme (p. 301):

"Heyting's primitive notions are to be translated as follows:

$\neg p$	$\sim Bp$
$p \supset q$	$Bp \rightarrow Bq$
$p \vee q$	$Bp \vee Bq$
$p \wedge q$	$p \,.\, q$."

For example, "¬*p*" in intuitionistic logic ("It is provable that not-*p*."), is translated as "~*Bp*" in classical logic ("It is not the case that it is provable that *p*."). Gödel's translation allows classical logic to say, what intuitionistic logic wants to say, by using the linguistic resources available within classical logic. In other words, Gödel attempted to *classicalize* intuitionistic logic.[12]

[10] For example, when *p* is the twin prime conjecture - that there are infinitely many twin primes, i.e. pairs of primes that differ by 2, such as {3, 5}, {5, 7}, {11, 13}, {17, 19}. Both the twin prime conjecture and its negation are still not provable.

[11] "*B*" for "beweisbar", the German word for "provable".

[12] Two most important questions, necessary for an assessment of Gödel's classicalization of intuitionistic logic, are, however, out of the scope of this thesis:

1. How *successful* is Gödel's translation?
2. How *classical* is the enriched classical logic?

Analogously, in this thesis, I attempt to *classicalize fuzzy logic*. I.e. I attempt to give an interpretation of fuzzy logic within classical logic enriched by degree operators {○, ◔, ◑, ◕, ●} - "it is of no/low/moderate/high/full degree that…". For example, when "*p*" has a 0.5 degree-of-truth according to fuzzy logicians, "*p*" will be translated as "◑*p*" ("it is of moderate degree that *p*.") in our enriched classical logic. If such classicalization of fuzzy logic is successful, then the charge that degree theories use non-classical logic will be undermined. Because degree theorists can still say what they want to say, by using the language of our enriched *classical* logic.

1.4 Methodology

We start with two less controversial claims:

1. the word "dark" is vague; and
2. **this** is darker than **this!**

Figure 1: A black-white colour spectrum

As figure 1 shows, *darkness comes in degrees*. That is to say, vague things - vague or really precise as they are - are *of degrees*. They are also *comparable*. Now *how things are of degrees* is what needs to be explained - our *explanandum*. The *explanans*, as it is, is *degrees*.

Degree-theorists have therefore tried hard to *make something of degrees* to be the *explanans*. Many claim that *truth* is of degrees.[13] Some claim that *predication* is of degrees.[14]

[13] E.g. Lakoff (1973), Zadeh (1975), and Sainsbury (1986). Machina (1976 p. 54) writes explicitly that his degree of truth "is *not* an *epistemic* notion like degrees of certainty".

[14] Forbes (1983 and 1985 p.170) argues for both degree of truth and degree of predication. In 1983 p. 241-242, Forbes says the following chain of inference is "hard to find a well-motivated objection".
Take two shades of red patches such that A is redder than B:

1. From A is redder than B, we can infer
2. A is red to a greater degree than B, we can infer
3. A satisfies the predicate "is red" to a greater degree than B, we can infer
4. The judgement that *A is red* has a higher degree of truth than the judgement that *B is red*.

3 is committed to degree of predication; 4 to degree of truth.

Some stop talking about truth-values, instead they talk about *epistemic statuses* like uncertainty, probabilities, and partial beliefs which *are* of degrees (Edgington 1992, 1997).

These seem to me the wrong steps taken by degree-theorists. Because when the *explanans* is degrees - i.e. when the degrees *themselves* are the only things that carry the entire explanatory weight; when the degrees are all that we need -, we should not rashly *fuse the degrees into truth*; we should not rashly fuse the degrees into predication; we should not give up or ignore bivalent truth-values just to borrow the degrees that we need from epistemic statuses. Instead, *we should simply propose degrees as something independent first*, let degrees do all its explanatory works, leave everything else untouched, and see how it goes. Here we have conceptually separated degrees from truth(/falsity).

In chapter 4 we will formally separate degrees from truth(/falsity) in our semantics. In short, *degree-of-truth is avoided; but degrees and truth are preserved*. The notion of degree-of-truth will be classicalized as classical bivalent truth-value (T/F), and a largely independent notion of degrees. Pictorially, °T is broken down to ° and T.

The separation of degree-of-truth into degrees and truth(/falsity) is an improvement upon degree theory. It is an improvement in this sense: *we trade some unifiedness of degree theory for the preservation of classical logic*. Degree-of-truth is broken down to degrees and truth, so some unifiedness of our theory is lost. (Note, however, that degree-of-truth is itself a very controversial notion.[15]) No *simplicity* is lost, because both degrees and truth are present in the theory before and after their separation. But what is gained is the preservation of classical logic. This is a trade that a lot of philosophers are happy to make.

1.5 Outline

This thesis does not contain a solution to the Sorites paradox. This thesis aims to contribute to the discussions of the semantics of degree, and the *(in)validity* of the Sorites paradox. The main outcomes of this thesis are 1. a semantics of degree, which is an enriched classical logic, which is then used to provide 2. an analysis of the *(in)validity* of the Sorites

[15] Frege, for example, thinks that what is true does not admit of degree:

> "We can see, to begin with, that what is beautiful admits of degrees, but what is true does not. We can think two objects beautiful, and yet think one more beautiful than the other. On the other hand, if two thoughts are true, one is not more true than the other." - Frege, *Logic* (1897), in eds. Beaney 1997 p. 231–232.

argument. I hope the new semantics will be compelling to both classical and non-classical logicians.

Chapter 2 informally introduces our degree operators. Chapter 3 justifies the application of degree operators to complex sentences. Chapter 4 presents our formal syntax and semantics. Chapter 5 derives two semantic consequences from our semantics. Chapter 6 applies the two semantic consequences to analyse the (in)validity of the Sorites argument. There are two results: 1. the validity of the standard Sorites argument is reasserted, and 2. a new argument for the *in*validity of the degreed version of the Sorites argument is presented. Each chapter ends with a brief summary of that chapter.

Summary of Chapter 1

In chapter 1.1, I have presented the Sorites paradox. In chapter 1.2, I have presented the standard degree-of-truth theory, its solution to the Sorites paradox, and the cost of that solution. In chapter 1.3, I have presented Gödel's classicalization of intuitionistic logic. In chapter 1.4, I have presented our methodology. In chapter 1.5, I have presented an outline of the following chapters.

Chapter 2. Introducing Degree Operators

Adverbs of degree can be put into 5 groups of near-synonyms:

1. not at all ...
2. barely, just, quite, scarcely, somewhat ...
3. fairly ...
4. largely, greatly, highly, intensely, mostly, very …
5. completely, entirely, fully, thoroughly, totally …

The first thing to notice is that adverbs of degree have a descriptive use that affect the truth-values of the sentences containing them. Consider the sentences:

(1) Ayer is a good philosopher.

(2) Ayer is a very good philosopher.

The picture here is that there are philosophers, of which only some are good, of which only some are very good. {x: x is a very good philosopher} is a proper subset of {x: x is a good philosopher}. So (2) entails (1), but (1) does not entail (2). Semantically, the word "very" in "... very good philosopher" may be formalized as a function that takes the extension of the expression from the set of good philosophers to the set of very good philosophers.

Suppose Ayer is only a fairly good philosopher, then (1) is true but (2) is false. So (1) and (2) can have different truth-values. Since (1) and (2) can have different truth-values under the same circumstance of evaluation, we can conclude that (1) and (2) have different meanings. Since the only linguistic difference between (1) and (2) is the absence and presence of the word "very", we can conclude that the difference in meaning between (1) and (2), and thereby their difference in truth-value under the same circumstance of evaluation, is due to the absence and presence of "very". The word "very" in (2) contributes to the overall truth-related meaning of (2). To use Ayer's word, "very" is not "cognitively meaningless", i.e. "very" affects truth-value. So are other adverbs of degree.

Now since (1) and (2) have different meanings and are *not* logically equivalent, if (1) is symbolized by "*p*", then (2) cannot be fully symbolized by the same logical symbol "*p*". That is to say, a logical symbol for "very" is wanted. So are other adverbs of degree. So I propose to *take adverbs of degree as operators*.

Corresponding to the 5 groups of adverbs of degree, the operators that I wish to introduce are:

○: It is of no degree that …

◔: It is of low degree that …

◑: It is of moderate degree that …

◕: It is of high degree that …

●: It is of full degree that ...[16]

[16] Two points to note:

1. Degree operators, like their English counterparts, can be rendered vague or sharp.
2. For simplicity I only propose five degree operators, and will only use one ("◕"). But we could propose more.

I propose to analyse degreed sentence by their degree-part and their sentence-part. For example, “Frege is a very good philosopher.” will be formalized as:

“It is of high degree that …” + “Frege is a good philosopher.”,
which we will formalize as “◕*p*”.

“I can barely play piano.” will be formalized as:

“It is of low degree that …” + “I can play piano.”,
which we may formalize as “◔*q*”.

The merit is *compositionality*, and with it a *clear bottom-up construction of truth-condition*. E.g. the sentence “Frege is a very good philosopher.” is true iff it is of high degree that Frege is a good philosopher. “I can barely play piano.” is true iff it is of low degree that I can play piano.

Our introduction of degree operators is a conservative extension of classical logic. The result is a more expressive language that is 1. classical, and 2. has the arsenal to talk about the degrees of sentences.

Summary of Chapter 2

In chapter 2 I have proposed to take adverbs of degree as operators. I have proposed to analyse degreed sentence by their degree-part and sentence-part.

Chapter 3. The Degrees of Sentences

Now, some readers may think it is okay to say that the sentence “Frege is a very good philosopher.” is of degree. After all, that sentence has an adverb of degree (“very”) that modifies a *gradable adjective* (“good”). However, it does not seem that other atomic sentences (such as “10,000 grains of sand make a heap.”), or complex sentences (“If 10,000 grains of sand make a heap, then 9,999 grains of sand make a heap.”), also has a degree component in a non-vacuous way. In short, the worries are:

1. Adverbs of degree are adverbs that typically make modification to *gradable adjectives*, not to entire sentences, and still less to complex sentences.
2. It is particularly perplexing to talk about the degree of conditional sentence. After all, we do not say things like "It is very if A then B." It may even be a category mistake to apply adverb of degrees to conditional, like predicating "sleeps" to "idea". In any case, "◕(A→B)" looks at best undefined.

In this chapter, I will justify the application of degree operators to all atomic sentences, their negations, conjunctions, and conditionals. I.e. I will make sense of the formulas "◕φ", "◕~φ", "◕(φ & ψ)", and "◕(φ→ψ)". I will argue that all atomic sentences and their logical compounds are degree-applicable.

3.1 The Degree of Atomic Sentence

First of all, consider atomic sentences that do not explicitly have a gradable adjective. Consider the following scenarios:

a. A book is on the fringe of a table. Half of the book is on the table, half of it is not. Its center of mass is on the very edge of the table. Now someone who cannot see the table asks you: "Is the book on the table?" It makes sense to say: "Well, the book is somewhat on the table." It also makes sense to say "The book is about 50% on the table."
b. An old man is jogging very slowly. Now someone who cannot see the old man asks you: "Is the old man running?" It makes sense to say: "The old man is barely running."

   We can also make *comparisons of degrees*. Suppose a second old man is jogging behind the first one, even more slowly. "Is the *first* man running?" It makes sense to say in response: "Well, he is to a greater degree running than the second old man." That's about as informative an answer as you can give when you don't know whether either man is running.
c. A glass of cocktail is on the bar table. 75% volume of the cocktail is apple juice, the rest are honey, alcohol, etc. "Is the cocktail apple juice?" It makes sense to

say: “The cocktail is largely apple juice.” It also makes sense to say “The cocktail is 75% apple juice.”

We can also make comparison: “Is this apple juice?” “It is to a greater degree an apple juice than the one you drank yesterday. In fact, this cocktail is to the degree of 3/4ths an apple juice.”

d. There is a liger in the zoo. Interbreeding occurs naturally. Liger is the hybrid offspring of male lion and female tiger. A child points at the liger, and asks you: “Is that a tiger?” It makes sense to say: “To a certain extent, it is a tiger.” It also makes sense to say: “It is partly a tiger.”, “It is a 50% tiger.”, “It is not entirely a tight, it is also not entire not a tiger.”

   We can also make comparison: “It is more tiger than that [point to its lion father].”

Notice that none of the sentences:

a. “The book is on the table.”
b. “The old man is running.”
c. “The cocktail is apple juice.”
d. “It is a tiger.”

contains a gradable adjective. Yet they are all perfectly degree-applicable. The point is this: *the source of degree of a sentence can be non-adjective*. Most *verbs* are of degree e.g. “love” (how *deep* is your love), or “run”, due to some roughly quantifiable scale, or some behavioural norm, from which behaviours may deviate. Some people run like more jogging, some people run more like sprinting. Most *nouns* are of degree, or could easily be, e.g. “Bernard Williams”, “the ship of Theseus”, because of the referent’s tolerant identity and (changing) physical composition. Collective nouns like “cat” are of degree, as mutations give rise to individual differences, and interbreeding gives rise to hybrid and sub-species. Many *prepositions* are of degree, e.g. “the fat man is *in* the room.” (when only half of his body is inside the door), due to more, or less, spatial occupation.

How about logical and mathematical statements? Answer: their degrees are always either 1 or 0. For our semantics to work, every sentence has to apply to a degree in [0,1]. This is compatible with some applying only to a degree in {0,1}. Their degrees are either 0 or 1 because of sharply stipulated definitions.

However, it seems safe to say that most ordinary English sentences have a degree-value that falls *between* 0 and 1, although normally we are not aware that it is so, because a precision in degree is not what we are usually concerned with, nor do we usually demand, when we talk to each other. But we could, always, ask: "Okay, but, to what degree?".

Now, since we are *not* going to eliminate - instead we want to capture - the degree-components in English sentences, when we translate them into our new logic, so *all* atomic sentences in our new logic will have a degree-value. *A fortiori*, all atomic sentences in our logic will be degree-applicable, despite the absence of a gradable adjective in their original English sentences. So it is justifiable to apply degree operators to entire sentences, and write "◕φ".

3.2 The Degree of Negation

When all atomic sentences are degree-applicable, their negations look as degree-applicable. Consider e.g. "The book is somewhat *not* on the table." So it makes sense to apply degree operators to negations, and write "◕~φ".

3.3 The Degree of Conjunction

Now suppose there are two books on the table, one somewhat on the table, one fully on the table. The sentences "This book is somewhat on the table." and "That book is fully on the table." both have a degree-component. Now consider the conjunction of the atomics: "This book is somewhat on the table and that book is fully on the table." Question: does the entire conjunction have a degree-component?

My claim is modest: it makes sense to construct *a* logical conjunction that is degree-applicable, and is *degree-functional*. That is, it makes sense to say that there *is* a degree for such conjunction, and that the degree of such conjunction is completely determined by the degrees of its conjuncts.

There are a number of reasonable ways to calculate the degree of conjunction from the degrees of its conjuncts. Let's write the degree of A as "°A". We may, for example, define °(A

& B) as (°A + °B) / 2, i.e. the degree of a conjunction is the average degree of its conjuncts's. We may define °(A & B) as min. {°A, °B}, i.e. the degree of a conjunction is the minimal degree of its conjuncts's. Here we do not need to choose. And they are simply two different conjunctions. My point here is that it makes sense to construct a logical conjunction that is degree-applicable, and that the conjunction's degree is determined by the degrees of its conjuncts. So it is justifiable to apply degree operators to conjunctions, and write "◕(φ & ψ)".

3.4 The Degree of Conditional

We have argued that "~" and "&" are both degree-applicable. Now since {~, &} are truth-functionally complete, we can define "→" in terms of "~" and "&". (I.e. "(φ→ψ)" can be defined as "~(φ & ~ψ)".) But then conditionals are degree-applicable. For if two formulas are inter-defined, the degree-applicability of one is the same degree-applicability of the other. I do not see how we can apply degree operators to most atomic sentences, to their negations, to their conjunctions, but not to their *inter-definable* conditionals.

In this footnote[17] I show how to *paraphrase conditional as a gradable adjective*. All in all, it seems safe to say that conditionals are degree-applicable, so it makes sense to write "◕(φ→ψ)".

Summary of Chapter 3

In chapter 3, I have justified the application of degree operators to all atomic sentences, their negations, conjunctions, and conditionals. In other words, I have provided justifications for writing "◕φ", "◕~φ", "◕(φ & ψ)", and "◕(φ→ψ)", which we will write in chapters 4 to 6.

Chapter 4. Degree Operators Formally

[17] Unofficially, we can understand conditional as an adjective, and let adverbs of degree modify it. (In fact, "conditional" is both a noun and an adjective.) "If A then B." can be understood as "B is conditional upon A." I.e. the conditional [noun] "If A then B." can be understood as "B is conditional [adjective] upon A.". "A→" can then be understood as a predicate of "B".
(We should say "B is very *materially* conditional upon A.", if we want to make sure that we are talking about material conditional.)
While it sounds exotic to say "It is *very* if A then B", it sounds perfectly sensible to say "B is *very* conditional upon A.", and "C is only *barely* conditional upon A." For example, my PhD admission is very conditional upon the quality of my writing sample, but my PhD admission is only barely conditional upon my GRE result.

In chapter 2, we have justified the introduction of degree operators. In chapter 3, we have justified the application of degree operators to all atomic sentences and their logical compounds. Now let me present our formal syntax and semantics for degree operators.[18]

Preserve everything in standard propositional logic, and add the following syntax and semantics.

4.1 The Syntax of Degree Operators

Formation rule:

If "φ" is a wff, then "○φ", "◔φ", "◑φ", "◕φ", and "●φ" are also wffs.

4.2 The Semantics of Degree Operators

A model **M** is an ordered pair ⟨*v*, *d*⟩ - valuation and degree - where *v* is a function from the atomic sentences to {T, F}, and *d* is a function from the atomic sentences to the real numbers in [0, 1].[19]

The semantic value of any wff φ in M, written "$[[\varphi]]_M$", is:

$$[[\varphi]]_M = \langle\, v(\varphi), d(\varphi)\,\rangle$$

For example: Frege is such a good philosopher. Suppose it is to degree 0.99 that Frege is a good philosopher. Express "Frege is a good philosopher." by "*p*", then $v(p) = T$, $d(p) = 0.99$, and $[[p]]_M = \langle\, T, 0.99\,\rangle$.

Next, we take "~" and "→" as our primitive logical connectives. Degree-functionality - that the degree of complex sentence is determined by the degrees of its atomic sentences - enables a systematic treatment of the degrees of formulas. Other things being equal, we want degree-functionality. To get degree-functionality, the degrees of "~φ" and of "(φ → ψ)" have to be functions of the degrees of "φ" and of "ψ". Given the degrees of "φ" and "ψ", what should be the degrees of "~(φ)" and "(φ → ψ)"?[20]

---

[18] This is *not* a full presentation of degree logic. The logical details are minimized for the purpose of this thesis, which is to introduce a new semantics to analyse the *(in)validity* of the Sorites argument.

[19] The "denseness" of real number may capture "higher-order vagueness". That is, that there is a real number between any two real numbers, may capture, that there is a borderline case between any two borderline cases.

[20] To improve readability, sometimes "A" or "x" is substituted for "φ"; and "B" or "y" for "ψ".

4.2.1 $d(\sim\varphi)$

Negation first. If we are to choose an equation for negation to govern the *degree-of-truth* of formula, there are only two plausible options: Łukasiewicz negation, and Gödel negation:

Łukasiewicz negation $\sim_Ł$

$°T(\sim_Ł\varphi) = 1 - °T(\varphi)$

That is, the degree-of-truth of a negated formula is (1 - the degree-of-truth of the formula that is being negated). For example, if *p* is true to degree 0.3, then $\sim_Ł p$ is true to degree (1 - 0.3) = 0.7.[21]k

Gödel negation $\sim_G$

$°T(\sim_G\varphi) = 1$ iff $°T(\varphi) = 0$

$°T(\sim_G\varphi) = 0$ iff $°T(\varphi) > 0$

That is, the degree-of-truth of a negated formula is 1 iff the degree-of-truth of the formula that is being negated is 0. The degree-of-truth of a negated formula is 0 iff the degree-of-truth of the formula that is being negated is larger than 0. Since the degree-of-truth of any formula is between 0 and 1 both inclusive, so the degree-of-truth of any formula is either 0 or > 0, so the degree-of-truth of any negated formula is either 1 or 0. Intuitively, the idea is that although there are degrees-of-truth, there are no degrees-of-falsity. Falsity is always "full" falsity. When a formula is true, to however low a degree, so long as its degree-of-truth is larger than 0, saying that "its negation is fully false." is (fully) false.

However, in the case of *degree*, I think there are only two plausible options for negation:

Generalized Łukasiewicz negation $\sim_Ł$

$d(\sim_Ł\varphi) = 1 - d(\varphi)$

[21] Łukasiewicz negation is proposed in Łukasiewicz and Tarski 1930. In Theorem 16 we read "g (x) = 1 – x". Łukasiewicz and Tarski were speaking in terms of countably infinitely many *truth-values*.

That is, the degree of a negated formula is (1 - the degree of the formula that is being negated).

A degree-inert negation $\sim_{=}$

$d(\sim_{=}\varphi) = d(\varphi)$

That is, the degree of a negated formula is same as the degree of the formula that is being negated. What this degree-inert negation does is that 1. truth-value-wise, it flips the truth-value of the formula that is being negated from T to F, or from F to T, and 2. degree-value-wise, it does nothing but to preserve the degree-value of the formula that is being negated.[22]

We do *not* need to choose one of these negations, since they are just two different negations. And we will not do any negation in this thesis, since there is no negation operation being done in the standard Sorites paradox.

4.2.2 $d(\varphi \rightarrow \psi)$

Next, conditional or material-implication. If we are to choose an equation to govern the *degree-of-truth* of material implication, then we have three more plausible and popular options:

Łukasiewicz triangular-norm, which gives "Łukasiewicz implication" $\rightarrow_{Ł}$[23]

$°T(x \rightarrow_{Ł} y) = \min \{(1 - x + y), 1\}$

Gödel triangular-norm, which gives "Kleene-Dienes implication" $\rightarrow_{G}$

$°T(x \rightarrow_{G} y) = \max \{(1 - x), y\}$

---

[22] Generalized Gödel negation $\sim_{G}$

$d(\sim_{G}\varphi) = 1$ iff $d(\varphi) = 0$

$d(\sim_{G}\varphi) = 0$ iff $d(\varphi) > 0$

Generalized Gödel negation $\sim_{G}$ is ruled out because it makes no intuitive sense: there is no consideration like "all falsities are full falsities" for the negation operation of *degree*. It seems hardly justifiable that if $d(\varphi) > 0$, then $d(\sim_{G}\varphi) = 0$. Worse still, when $d(\varphi) > 0$, generalized Gödel negation *loses track of the exact degree of the atomic sentence*. As a result $d(\sim_{G}\sim_{G}\varphi) \neq d(\varphi)$. This lost of information in negation operation is undesirable. Also, we may want double negation to be both truth-value and degree-value inert, i.e. double negation does not change the truth-value nor the degree-value of the formula that is being doubly negated.

[23] Łukasiewicz implication is proposed in the same Łukasiewicz and Tarski 1930. In the same Theorem 16 we read "f (x, y) = min. (1, 1 – x + y)". Again, Łukasiewicz and Tarski were speaking in terms of countably infinitely many *truth-values*.

Product/Goguen triangular-norm, which gives "Reichenbach implication" $\rightarrow_P$

$^{\circ}T(x \rightarrow_P y) = 1 - x + x \cdot y$

The mathematical properties of these triangular-norms are well-studied e.g. in Klement, Mesiar, and Pap's *Triangular Norms* (2000).[24]

In the case of the *degree* of conditional, we have to make a decision. The minimal decision that we have to make is about which implication to use to formalize the second to the last premisses of the standard Sorites paradox.[25] Let's generalize the three triangular-norms to the case of degree first:

Generalized Łukasiewicz triangular-norm

$d(x \rightarrow_Ł y)$	=	1	iff	$d(x) \leq d(y)$
		$1 - d(x) + d(y)$	iff	$d(x) > d(y)$

Generalized Gödel triangular-norm

$d(x \rightarrow_G y)$	=	1	iff	$d(x) \leq d(y)$
		$d(y)$	iff	$d(x) > d(y)$

Generalized Product/Goguen triangular-norm

$d(x \rightarrow_P y)$	=	1	iff	$d(x) \leq d(y)$
		$d(y)/d(x)$	iff	$d(x) > d(y)$

Which generalized triangular-norm, if any of the above three, should we choose, for the degree of conditional, whose degree-value is to be degree-functionally determined by the degree-values of the antecedent and of the consequent? The approach that we will take is to start with some very plausible assumptions about the behaviour of the degree of conditional, then

[24] Especially in chapter 11.1, all three triangular-norms can be seen on p. 235. For an easy introduction see Priest 2008 chapter 11.7a "*Appendix: *t*-norm Logics" p.234–237.

[25] "Premiss 2: If 10,000 grains of sand make a heap, then 9,999 grains of sand make a heap.
Premiss 3: If 9,999 grains of sand make a heap, then 9,998 grains of sand make a heap.
…
Premiss 10,000: If 2 grains of sand make a heap, then 1 grain of sand makes a heap.
Premiss 10,001: If 1 grain of sand makes a heap, then 0 grain of sand makes a heap."

look at a few cases, and then come up with the simplest equation that satisfies all those assumptions and what we want to say in those cases.

Some very plausible assumptions:

1. $0 \leq d(x \rightarrow y) \leq 1$

   The degree of conditional is within 0 to 1, both inclusive.
2. $d(x \rightarrow x) = 1$

   The degree of self-implication is full.
3. If $d(x) = d(x^*)$, then $d(x \rightarrow y) = d(x^* \rightarrow y)$

   If $d(y) = d(y^*)$, then $d(x \rightarrow y) = d(x \rightarrow y^*)$

   The substitution of the antecedent or of the consequent, with a formula that has the same degree, does not change the degree-value of the entire material implication.[26]
4. If $d(y) < d(y^*)$, then $d(x \rightarrow y) \leq d(x \rightarrow y^*)$

   If the degree of the antecedent remains the same, increasing the degree of the consequent increases the degree of the entire conditional. Unless the degree of the original conditional is already 1, then the degree of the resulting conditional is also 1. Put it in a different way: increasing the degree of the consequent cannot lower the degree of the entire conditional.

Mathematically, in $(x \rightarrow y)$, either:

A. $d(x) = d(y)$, or
B. $d(x) < d(y)$, or
C. $d(x) > d(y)$

A. If $d(x) = d(y)$, then, by assumptions 2 and 3, $d(x \rightarrow y) = 1$.
B. If $d(x) < d(y)$, then, by assumptions 2 and 4, $d(x \rightarrow y) = 1$.
C. If $d(x) > d(y)$, then, we have a real decision to make. Let's look at a few cases:

[26] The *truth*-value of material implication is only interested in the truth-values, but not the degree-values or the contents, of the antecedent and of the consequent. Likewise, the *degree*-value of *material* implication is only interested in the degree-values, but not the truth-values or the contents, of the antecedent and of the consequent.

	d(x)	d(y)	d(x→y)
I	0.9	0.7	?
II.	0.8	0.7	?
III.	0.5	0.1	?
IV.	0.4	0	?

What should we make to be the degree-value of d(x→y) in these cases?

Case I: a reasonable answer for d(x→y) would be 0.7. Because 0.7 is the degree that is being preserved in the consequent. This, in fact, is what generalized Gödel triangular-norm issues: when $d(x) > d(y)$, then $d(x \to y) = d(y)$. But a quick comparison shows that something is missing in this answer: it ignores the degree of x completely, so it cannot capture any numerical distance or relation between d(x) and d(y), except that $d(x) > d(y)$.

Consider case I together with case II: both consequents are of degree 0.7. Do we want to say that, as Gödel triangular-norm suggests, that both material implications are therefore of the same degree 0.7? Not really. This is one reason why: because in case II, *a larger proportion of the degree of the antecedent* is preserved after the material implication operation. So, in order to indicate the proportion of the degree of the antecedent being preserved in the consequent, a more informative value of d(x→y) for case I would be, e.g. 0.7/0.9 = 0.777…, and for case II 0.7/0.8 = 0.875. Now 0.875 > 0.777, so the degree of the material implication in case II is higher, and this indicates that a larger proportion of the degree of the antecedent is preserved in the consequent after the material implication operation in case II. This, in fact, is what generalized Product/Goguen triangular-norm issues: when $d(x) > d(y)$, then $d(x \to y) = d(y)/d(x)$.

The problem with generalized Product/Goguen triangular-norm is that we are dealing with *material* implication, the antecedent of which can be totally irrelevant in content to the consequent. Although the degree of the consequent is lower than the antecedent's, the degree of the consequent need not be a part or proportion of the degree of the antecedent. We want a simpler relation between d(x) and d(y) than division and what division may suggest. So, what other options do we have?

Mathematically there are too many options. To choose one of the too many options, we must ask ourselves one question: in (x→y), when the degree of the antecedent is higher than the

consequent's, $d(x) > d(y)$, *what is the thing that we want most to capture, with the degree of the material implication?* The answer, I think, is this: *the amount of degree dropped/lost* during the material implication operation.

This is very simple. Consider case I, the degree of the antecedent is 0.9. The degree of the consequent is 0.7. Question: degree-value-wise, what did the material implication do onto the 0.9 degree of the antecedent, to make the 0.9 to become 0.7? Well, - 0.2!

When $d(x) > d(y)$, since *the amount of degree dropped* from the antecedent to the consequent is what we want most to capture with the degree of material implication operation, the simplest move is to make the amount of degree dropped *to be* the degree of the material implication. In case I, since what the material implication did to the degree of the antecedent is to - 0.2, we can most straightforwardly capture that amount of degree dropped by saying that the degree of that material implication *is* -0.2. In case II, $d(x \rightarrow y) = (0.7 - 0.8) = -0.1$. In case III, $d(x \rightarrow y) = (0.1 - 0.5) = -0.4$. In case IV, $d(x \rightarrow y) = (0 - 0.4) = -0.4$. All well and good.

One complication: according to our semantics or very plausible assumption 1, the degree-value of material conditional has to be within 0 and 1 both inclusive. But we've just proposed to make the degrees of material implication in case I to IV to be -0.2, -0.1, -0.4, and -0.4. How can we make the suggested degrees of material implication fall within 0 and 1 both inclusive? Well, simply + 1 after $d(y) - d(x)$! Since $d(x) > d(y)$, and both $d(x)$ and $d(y)$ are within 0 and 1 both inclusive, $1 + d(y) - d(x)$ must be within 0 and 1 both inclusive. The resulting number still tracks the amount of degree dropped during the material implication.

So we got our answer: when $d(x) > d(y)$, $d(x \rightarrow y) = 1 + d(y) - d(x)$. This, in fact, is what generalized Łukasiewicz triangular-norm issues: $d(x \rightarrow_{Ł} y) = \min\{(1 + y - x), 1\}$. So we will use generalized Łukasiewicz triangular-norm to calculate the degree of conditional from the degrees of the antecedent and of the consequent. We may rewrite it:

Generalized Łukasiewicz triangular-norm

$d(A \rightarrow_{Ł} B) = 1 + \min\{ d(A), d(B) \} - d(A)$

Thus

$d(A \rightarrow B) = 1$ iff $d(\text{consequent}) \geq d(\text{antecedent})$

$d(A \rightarrow B) = 1 + d(B) - d(A)$ iff $d(\text{antecedent}) > d(\text{consequent})$

Unlike the case of negation where we did not need to choose, we are committed to generalized Łukasiewicz triangular-norm, and we will use it to derive our results in chapters 5 and 6. Gödel and Product/Guoguan triangular-norms do *not* produce our results.

4.2.3 d(A & B) and d(A ∨ B)

{~, →} is truth-functionally complete, i.e. by using {~, →} we can define all the other binary bivalent truth-functional operators, e.g. ∨, &, and ↔. However, truth-functional completeness is independent of degree-functional completeness. If we are to make {~, →} *degree*-functionally complete, then we will have to define all the other binary *degree*-functional operators by using {~, →}. This can be done, but the result may not be what we want.

Suppose we use generalized Łukasiewicz negation and generalized Łukasiewicz triangular-norm:

Generalized Łukasiewicz negation $\sim_Ł$

$d(\sim_Ł \varphi) = 1 - d(\varphi)$

Generalized Łukasiewicz triangular-norm

$d(A \rightarrow_Ł B) = 1 + \min\{ d(A), d(B) \} - d(A)$

Then, just like we can define (A & B) to be ~(A → ~B), we can define d(A & B) to be d~(A → ~B).[27]

$$
\begin{aligned}
d\sim(A \rightarrow \sim B) \quad = \quad & 1 - d(A \rightarrow \sim B) \\
& 1 - [1 + \min\{ d(A), d(\sim B) \} - d(A)] \\
& 1 - [1 + \min\{ d(A), [1 - d(B)] \} - d(A)]
\end{aligned}
$$

So, $d(A \&_Ł B) = 1 - [1 + \min\{ d(A), [1 - d(B)] \} - d(A)]$

[27] More precisely, we can define $d(A \&_Ł B)$ to be $d\sim_Ł(A \rightarrow_Ł \sim_Ł B)$.

Similarly, just like we can define (A ∨ B) to be (~A → B), we can define d(A ∨ B) to be d(~A → B).[28]

$$d(\sim A \rightarrow B) = 1 + \min\{ d(\sim A), d(B) \} - d(\sim A)$$
$$1 + \min\{ [1 - d(A)], d(B) \} - [1 - d(A)]$$

So, $d(A \vee_{Ł} B) = 1 + \min\{ [1 - d(A)], d(B) \} - [1 - d(A)]$

Now suppose d(A) = 0.7, and d(B) = 0.4, then accordingly:

$$\begin{aligned} d(A \&_{Ł} B) &= 1 - [1 + \min\{ d(A), [1 - d(B)] \} - d(A)] \\ &= 1 - [1 + \min\{ 0.7, [1 - 0.4]\} - 0.7] \\ &= 1 - [1 + \min \{0.7, 0.6\} - 0.7] \\ &= 1 - [1 + 0.6 - 0.7] \\ &= 1 - 0.9 \\ &= 0.1 \end{aligned}$$

$$\begin{aligned} d(A \vee_{Ł} B) &= 1 + \min\{ [1 - d(A)], d(B) \} - [1 - d(A)] \\ &= 1 + \min\{ [1 - 0.7], 0.4\} - [1 - 0.7] \\ &= 1 + \min \{ 0.3, 0.4 \} - 0.3 \\ &= 1 + 0.3 - 0.3 \\ &= 1 \end{aligned}$$

So, if we define conjunction and disjunction by $\{\sim_{Ł}, \rightarrow_{Ł}\}$, then when d(A) = 0.7 and d(B) = 0.4, $d(A \&_{Ł} B) = 0.1$ and $d(A \vee_{Ł} B) = 1$. These results are odd. The much more plausible options are:

$$d(A \& B) = \min \{ d(A), d(B) \}$$
$$d(A \vee B) = \max \{ d(A), d(B) \}$$

[28] More precisely, we can define $d(A \vee_{Ł} B)$ to be $d(\sim_{Ł} A \rightarrow_{Ł} B)$.

That is, the degree of a conjunction is the degree of the lowest degree conjunct. The degree of a disjunction is the degree of the highest degree disjunct. $d(A \& B) = \min \{ d(A), d(B) \}$ is more plausible because we can argue that:

1. $d(A) \leq d(A \& A)$
   I.e. the degree of a self-conjunction is not lower than the degree of the self.
2. $d(A \& B) \leq d(A)$ and $d(A \& B) \leq d(B)$
   I.e. the degree of any conjunct is at least as high as the degree of its conjunction.
3. If $d(A) \leq d(A^*)$ and $d(B) \leq d(B^*)$, then $d(A \& B) \leq d(A^* \& B^*)$
   I.e. if each conjunct is substituted by a wff whose degree is as least as high, then the conjunction after substitution has a degree that is as least as high as before.

From 1-3 we can prove that $d(A \& B) = \min \{ d(A), d(B) \}$.[29]

For $d(A \vee B) = \max \{ d(A), d(B) \}$, sub "∨" to "&" in 1-3, and sub "≥" to "≤", then we can prove that $d(A \vee B) = \max \{ d(A), d(B) \}$. Or we can generalize De Morgan's Law $(A \vee B) =_{def} \sim(\sim A \& \sim B)$ to become $d(A \vee B) =_{def} d[\sim(\sim A \& \sim B)]$. Given $d(\sim A) = 1 - d(A)$, and $d(A \& B) = \min \{ d(A), d(B) \}$:

$$
\begin{aligned}
d[\sim(\sim A \& \sim B)] &= 1 - [d(\sim A \& \sim B)] \\
&= 1 - \min \{ d(\sim A), d(\sim B) \} \\
&= 1 - \min \{ [1 - d(A)], [1 - d(\sim B)] \} \\
&= \max \{ d(A), d(B) \}
\end{aligned}
$$

$$
\text{So } d(A \vee B) = \max \{ d(A), d(B) \}
$$

We do *not* need to choose one of these conjunctions or disjunctions, since & and $\&_Ł$, and ∨ and $\vee_Ł$, are just different conjunctions and disjunctions. Also, we will not do any conjunction or disjunction in this thesis, since there is no conjunction or disjunction operation being done in the standard Sorites paradox.

[29] The same argument and proof for the *degree-of-truth* of conjunction is presented in Williamson 1994 chapter 4.8 "Continuum-Valued Logic: Truth-Tables" p. 114–120, where °T(A & B) = min {°T(A), °T(B)}.

4.3 Tarskian Model-Theoretic Definition of Validity

Then we define validity. Following the standard Tarskian model-theoretic definition of logical consequence in a formalized language (Tarski 1936), a premiss set $\Gamma \vDash \varphi$ iff:

In all models in which every member of Γ is true, φ is also true.[30]

4.4 $v(\varphi)$ and $d(\varphi)$

The relation between $v(\varphi)$ and $d(\varphi)$ - i.e. between truth-value and degree-value - is *very* intricate. Do we really want to say that, e.g. if $d(\varphi) = 1$ then $v(\varphi) = T$? How about if $d(\varphi) = 0$ then $v(\varphi) = F$? How about $v(\varphi) = T$ iff $d(\varphi) \geq 0.5$; and $v(\varphi) = F$ iff $d(\varphi) < 0.5$?[31]

The relation between e.g. $[[\varphi]]_M$ and $[[◕\varphi]]_M$ is also *very* intricate. There are many ways to define these relations, each way has some different costs and benefits. It is not clear which way is better.

To illustrate one such intricacies, let's say "VERY" is a sentential operator that ascribes a 0.9 degree to its operand, so that "VERY φ" has a degree 0.9. Now, what is the degree of "VERY VERY φ"? We have at least four more straightforward options:

A. 0.9 + [(1 - 0.9) x 0.9] = 0.99.
B. 0.9.
C. (0.9 + 0.9) / 2 = 0.9.
D. 0.9 x 0.9 = 0.81.

Which one should we choose? Also, do we want to allow utterances such as "Hegel is a barely fairly very good philosopher."? If we do then how are we going to calculate the degree of that

[30] The *standard* Tarskian model-theoretic definition of validity is deliberately chosen for its standardness. I take this to be an advantage of my thesis: whatever turns out to be valid or invalid, it is so according to the most standard definition of validity. Some degree theorists, e.g. Edgington (1992 p. 193) and Smith (2008 p. 222), in arguing that the Sorites paradox is or is not valid, first invent their own definitions of validity, which may be accused of changing the subject.

[31] What are degrees? Is degree *accuracy*? Is degree Carnap's *degree of confirmation*? Is degree Ramsey's *partial belief*? Is degree *probability*? Official answer: for my semantics to work, I am not committed to any particular theory of degree. This thesis is *not* an explication of the pre-theoretic concept of degree. This thesis is on the *semantics* of degree. A semantics of degree is about how degree-operation works, not about what degrees are.

sentence? If we don't then why not? For the purpose of this thesis, which is to analyse the (in)validity of the Sorites argument, these intricacies have to be skipped.

For our purpose, all we need about the relation between degree and truth is this:

$v(◕\varphi) = T$ iff $([v(\varphi) = T]$ & $[d(\varphi) = 0.99])$

That is, the truth-value of "◕φ" is true iff "φ" is true and has the degree 0.99. Intuitively, "◕φ" is true iff 1. "φ" is already true, and 2. the "◕" in "◕φ" denotes the correct degree of "φ".

Summary of Chapter 4

In chapter 4, I have presented a formal syntax and semantics of degree operators. For the purpose of deriving the two semantic consequences in chapter 5, which will be applied to analyse the (in)validity of the Sorites paradox in chapter 6, our minimal commitments are:

1. Generalized Łukasiewicz triangular-norm
   $d(A \rightarrow B) = 1 + \min\{ d(A), d(B) \} - d(A)$

2. A relation between truth-value and degree-value
   $v(◕\varphi) = T$ iff $([v(\varphi) = T]$ & $[d(\varphi) = 0.99])$

Chapter 5. The Logic of Degree Operators

In this chapter I will derive two semantic consequences from our semantics given in chapter 4. The two semantic consequences will be needed in chapter 6, for our final analysis of the (in)validity of the Sorites argument.

5.1 $\{A, (A \rightarrow B)\} \vDash B$

This is (the semantic part of) *modus ponens*. *Modus ponens* is a classically valid rule of inference, and is therefore a valid rule of inference in our extended classical logic. *Modus ponens* is valid because in all models **M** in which $v(A) = T$, and $v(A \rightarrow B) = T$, therein $v(B)$ is also $= T$. In other words, *modus ponens* is truth-preserving in all models.

However, *modus ponens* is not *degree*-preserving in all models. This is a consequence of the separation of truth-value and degree-value in our semantics: *that an argument is truth-preserving (in all models **M**) does not entail that it is also degree-preserving (in all models **M**).*

A *counter-model* for the preservation of degree in $\{A, (A \rightarrow B)\} \vDash B$:

$[[A]]_M = \langle T, 0.99 \rangle$

$[[B]]_M = \langle T, 0.97 \rangle$

$v(A) = T$, and $v(B) = T$, so $v(A \rightarrow B) = \underline{T}$

$d(A) = 0.99$, and $d(B) = 0.97$

Since $d(A \rightarrow B) = 1 + \min\{ d(A), d(B) \} - d(A) = 1 + 0.97 - 0.99 = 0.98$

so $d(A \rightarrow B) = \underline{0.98}$

So $[[(A \rightarrow B)]]_M = \langle \underline{T}, \underline{0.98} \rangle$

Now there are two premisses $\{A, (A \rightarrow B)\}$. The degree of "A" = 0.99. The degree of "(A→B)" = 0.98. So the premiss that has the lowest degree is "(A→B)" with degree 0.98. But the conclusion, "B", has degree 0.97. And 0.97 < 0.98. In our semantics, the degree of a conclusion, which is truth-preservingly derived by *modus ponens* from a premiss set, can be *lower* than the degree of the premiss that has the lowest degree!

Although B *is* a logical consequence of $\{A, (A \rightarrow B)\}$, the degree of B can be lower than the degree of A and the degree of (A→B). In other words, the degree of a logical consequence can be lower than the degree of each of its premisses. To sum up, *modus ponens* is truth-preserving in all models, but it is not degree-preserving in all models.

5.2 $\nvDash ([\blacklozenge(A \rightarrow B)] \rightarrow (\blacklozenge A \rightarrow \blacklozenge B))$

The weakest normal system of modal logic - system K - has a distribution axiom:

Distribution axiom in system K:

$$([\Box(A \rightarrow B)] \rightarrow (\Box A \rightarrow \Box B))$$

Now consider,

Distribution axiom for ◕:

$$([◕(A \to B)] \to (◕A \to ◕B))$$

Question: Is $\vDash ([◕(A \to B)] \to (◕A \to ◕B))$ true?

Answer: No. A *counter-model* for $\vDash ([◕(A \to B)] \to (◕A \to ◕B))$:

$[[A]]_M = \langle T, 0.99 \rangle$

$[[B]]_M = \langle T, 0.98 \rangle$

Since $v(◕\varphi) = T$ iff $([v(\varphi) = T]$ & $[d(\varphi) = 0.99])$

so $v(◕A) = T$, but $v(◕B) = F$

$v(◕A) = T$, but $v(◕B) = F$, so $v(◕A \to ◕B) = \underline{F}$

$d(A) = 0.99$, and $d(B) = 0.98$

Since $d(A \to B) = 1 + \min\{ d(A), d(B) \} - d(A) = 1 + 0.98 - 0.99 = 0.99$

so $d(A \to B) = 0.99$

$v(A) = T$, and $v(B) = T$, so $v(A \to B) = T$

$v(A \to B) = T$, and $d(A \to B) = 0.99$, so $v[◕(A \to B)] = \underline{T}$

$v[◕(A \to B)] = \underline{T}$, and $v(◕A \to ◕B) = \underline{F}$, so $v([◕(A \to B)] \to (◕A \to ◕B)) = F$

In short, $([◕(A \to B)] \to (◕A \to ◕B))$ is false when $[[A]]_M = \langle T, 0.99 \rangle$, and $[[B]]_M = \langle T, 0.98 \rangle$. Pictorially:

([◕(A→B)]	→	(◕A	→	◕B))
T	T	T	T	T
T	***F***	*T*	*F*	*F*
T	T	F	T	T
T	T	F	T	F
F	T	T	T	T
F	T	T	F	F
F	T	F	T	T
F	T	F	T	F

Therefore, ([◕(A→B)]→(◕A→◕B)) is not a logical truth. The distribution axiom for ◕ does not hold. The degree of → is not in every case distributive over the antecedent and consequent. ⊭ ([◕(A→B)]→(◕A→◕B)).

Summary of Chapter 5

To sum up, in this chapter we have derived two semantic consequences from our semantics:

1. *Modus ponens* is truth-preserving in all models, but it is not degree-preserving in all models.
2. ⊭ ([◕(A→B)]→(◕A→◕B))

Chapter 6. The Sorites Paradox Revisited

In this final chapter, I will use the two semantic consequences derived in chapter 5 to analyse the *(in)validity* of the Sorites argument. There are two results:

1. The validity of the standard Sorites argument is reasserted.
2. A new argument for the *in*validity of the degreed version of the Sorites argument is presented.

6.1 Reasserting the Validity of the Standard Sorites Argument

For clarity let us formalize the standard Sorites argument given in chapter 1.1. Let "H_n" read "*n* number grain of sand makes a heap". Below is a standard formalization of a standard Sorites argument:

Premiss 1:	$H_{10,000}$
Premiss 2:	$(H_{10,000} \rightarrow H_{9,999})$
Premiss 3:	$(H_{9,999} \rightarrow H_{9,998})$
…	
Premiss 10,000	$(H_2 \rightarrow H_1)$
Premiss 10,001:	$(H_1 \rightarrow H_0)$
Conclusion:	H_0
Rule of inference	10,000 instances of *modus ponens*.

We can now analyse the (in)validity of the Sorites argument with our new semantics. In chapter 5.1 we have seen that *modus ponens* 1. is truth-preserving in all models, but 2. is not degree-preserving in all models. This is to say:

1. By *only* using *modus ponens*, which is truth-preserving in all models, the Sorites argument is a *valid* argument, according to the standard Tarskian definition of validity (given in chapter 4.3). That is to say, the Sorites conclusion *is* a logical consequence of the Sorites premisses. That is to say, in all models in which all the Sorites premisses are true, in those models the Sorites conclusion is true too.

2. However**,** by using *a lot* of *modus ponens*, which is not degree-preserving in all models, the Sorites conclusion may have a degree that is *a lot* lower than the degree of each Sorties premiss. In particular:
    a. The degree of the Sorites conclusion can be very low, even when the degree of *every* Sorties premiss is very high. *A fortiori*, the degree of the Sorites conclusion can be *significantly* lower than the degree of the Sorites premiss that has the lowest degree.
    b. Given $d(A \rightarrow B) = 1 + \min\{ d(A), d(B) \} - d(A)$, if the degree of every Sorites premiss is very high (which they certainly seem to be), then the difference in degree between two consecutive premisses is very small. Then, after each instance of *modus ponens*, the drop of degree for each inference from $\{A, (A \rightarrow B)\}$ to B is *very* small. That is to say, the leakage of degree in Sorites reasoning is very slow and gradual.
        i. This slow and gradual leakage of degree may explain why, when most people walk through the Sorites reasoning, after each instance of *modus ponens* they become more and more *uncertain* about the conclusion of that instance of *modus ponens*.
        ii. This mathematically shows how, in the Sorites argument, little by little, little discrepancies accumulate to a wild discrepancy.

Once again, the moral is this: *do not confuse degree with truth/falsity*. Unreflected common sense certainly does not sharply distinguish between the two. Even philosophers do not sharply distinguish between the two.

"Looking at your Sorites argument, I am not *sure* whether 0 grain of sand makes a heap. But it certainly looks very much not so." This kind of response to the Sorites paradox is perfectly natural. We are inclined to accept all Sorites premisses - whose degrees are very high -, and reject the Sorites conclusion - whose degree is very low. We may partly explain this inclination with our semantics.

Perhaps when we are given the Sorites argument, we recognize that the *degree* of the Sorites conclusion is very low, but then we immediately jump to the opinion that, *therefore* the

Sorites conclusion is *false*. This inference, from the very low degree of φ, to the falsity of φ, is unconscious, unjustified, and perhaps fallacious.

Is the inference, from the very low degree of φ, to the falsity of φ, valid? In our system, whether that inference is valid is an open question. Our system allows 1. true sentence to have a very low degree, and 2. false sentence to have a very high degree. (Recall that our minimal commitment about the relation between degree and truth is $v(◕\varphi) = \mathrm{T}$ iff $([v(\varphi) = \mathrm{T}]\ \&\ [d(\varphi) = 0.99])$.)

Now, according to our system, the Sorites argument is valid. *If* all the Sorites premisses are true, it follows that the Sorites conclusion is also true. But the Sorites conclusion has a very low degree, be it (the Sorites conclusion) true or false. So the Sorites conclusion may be a sentence that is true, and that its (the sentence's) degree is very low, or is even 0.

Low degree truth and high degree falsity may be repugnant. Or maybe not. Maybe the repugnance is nothing but the trace of the unreflected parallelization of degree with truth(/falsity). Here we cannot say more without committing ourselves to a particular theory of degree and a particular theory of truth.

Perhaps when we are thinking about the Sorites argument, we recognize that the degree of each Sorites premiss is very high. After all, whether 10,000 grains of sand really make a heap, it certainly looks very much so. Whether 5,000 grains of sand really make a heap, it certainly looks very much conditional upon whether 5,001 grains of sand make one. Then, when we do *modus ponens* with the Sorites premisses, we expect their logical consequence to look also very much so - to have a high degree. But this is a mistake. Truth-preservation does not entail degree-preservation, given how our semantics works.

Some degree theorists may insist on interpreting our degrees simply as degrees-of-truth, and ignoring the first (left) tuple of our model **M**, where $[[\mathrm{A}]]_{\mathbf{M}} = \langle\, v(\mathrm{A}), d(\mathrm{A})\,\rangle$. They would say that the Sorites argument is invalid, but this comes at the extremely heavy cost of rejecting LEM, and the validity of *modus ponens*. As a less revisionary logic, our semantics fully preserves LEM, fully preserves the validity of *modus ponens*, and everything in classical logic.

6.1.1 Remark on the Sorites Paradox

Now the validity of the Sorites argument is reasserted. That is to say, we have eliminated one of the three apparently jointly exhaustive possible solutions to the Sorites paradox:

1. A premiss is false.
2. ~~The argument is invalid.~~
3. The conclusion is true.

That is to say, if we adopt the enriched classical logic proposed in this thesis, then, apparently, in the standard Sorites argument, either a premiss is false, or the conclusion is true.

If a Sorites premiss is false, then, apparently, what needs to be explained is 1. why we don't know which premiss is false, and 2. how the meaning of the word "heap", or our concept of heap, or the identity of heap, are so sharp.

If the Sorites conclusion is true, then - since for any Sorites conclusion there is another equally compelling Sorites argument for the negation of that very Sorites conclusion -, not *every* Sorites argument has a true conclusion (, so some Sorites argument has a false premiss), unless we accept a true contradiction, namely that the Sorites conclusion is both true and false.

6.2 A New Argument for the *In*validity of the Degreed Version of the Sorites Argument

Now, given our new semantics, according to which the validity of the standard Sorites argument is reasserted, what would be the best argument for the conclusion that the Sorites argument is *in*valid? What could the degree theorists say with our new semantics?

First of all, we need to put ourselves into the shoes of degree theorists, and look at the Sorites paradox from *their* perspective. When degree theorist Edgington (1992) evaluates the Sorites argument, she thinks that she is reasoning with *uncertainty*: she is not *certain* whether each Sorites premiss is true, although she is *confident* that each is. She would then assign, e.g. a high *confidence*, or *probability*, or *degree-of-truth* to each premiss. Instead of the classical bivalent truth-value of the Sorites premisses, she is primarily concerned with their *epistemic status*, or *truth-status*, both of which are *of degree*.

With our degree operators, we can capture and classicalize such degree-theoretic reading of the Sorites argument. We can do so by prefixing each Sorites premiss and the Sorites conclusion with "It is of very high degree that …". The result is "*the Sorites° argument*". Degree theorists may interpret this "degree" to be whatever they think fit: a primitive notion of degree,

or accuracy, degree of confirmation, certainty, confidence, partial belief, credence, probability, etc., just not degree-of-truth.

Premiss 1:	$\blacksquare H_{10,000}$
Premiss 2:	$\blacksquare(H_{10,000} \rightarrow H_{9,999})$
Premiss 3:	$\blacksquare(H_{9,999} \rightarrow H_{9,998})$
…	
Premiss 10,000	$\blacksquare(H_2 \rightarrow H_1)$
Premiss 10,001:	$\blacksquare(H_1 \rightarrow H_0)$
Conclusion:	$\blacksquare H_0$

"$\blacksquare H_{10,000}$" reads "It is of very high degree that 10,000 grains of sand make a heap.", or "10,000 grains of sand is very heapy.".

"$\blacksquare(H_{10,000} \rightarrow H_{9,999})$" reads "It is of very high degree that if 10,000 grains of sand make a heap, then 9,999 grains of sand make a heap.", or "whether 9,999 grains of sand make a heap is very conditional upon whether 10,000 grains of sand make one.".

Now *this* argument - the Sorites° argument - is *in*valid, because:

Premiss I:	If $([\blacksquare(A \rightarrow B)] \rightarrow (\blacksquare A \rightarrow \blacksquare B))$ is not a premiss, nor a logical truth, then the Sorites° argument is invalid.
Premiss II:	$([\blacksquare(A \rightarrow B)] \rightarrow (\blacksquare A \rightarrow \blacksquare B))$ is not a premiss, nor a logical truth.
Conclusion:	The Sorites° argument is invalid.

6.2.1 On Premiss I

In order to derive the Sorites° conclusion "$◕H_0$" from the Sorites° premisses, the distribution axiom ([◕(A→B)]→(◕A→◕B)), or anything logically equivalent, is a *necessary* (and sufficient) component. If and only if we have ([◕(A→B)]→(◕A→◕B)), we can derive "◕B" from {◕A, ◕(A→B)}:

Formula	Rule of inference	Dependency
1. ◕A	Assumption	1
2. ◕(A→B)	Assumption	2
3. ([◕(A→B)]→(◕A→◕B))	Distribution axiom for ◕	3
4. (◕A→◕B)	2,3 *modus ponens*	2,3
5. ◕B	1,4 *modus ponens*	1,2,3

6.2.2 On Premiss II

However, ([◕(A→B)]→(◕A→◕B)) is not one of the 10,001 Sorites° premisses. Furthermore, in chapter 5.2 we have proven that ⊭ ([◕(A→B)]→(◕A→◕B)). In other words, ([◕(A→B)]→(◕A→◕B)) is not a logical truth.

6.2.3 On Conclusion

So the Sorites° argument is *invalid*. However, the Sorites° argument is invalid *not* because *modus ponens* is invalid. The rejections of *modus ponens* and LEM were the mistakes of the *old* degree theorists. The *new* degree theorists use our semantics to argue that the Sorites° argument is invalid because there is a missing component, whose presence is necessary for the syntactic derivation and semantic entailment, from the Sorites° premiss set to the Sorites° conclusion. The component is missing because it is not a premiss nor a logical truth. The missing component is the distribution axiom for ◕: ([◕(A→B)]→(◕A→◕B)).

Summary/Abstract

Since 1970s, degree-of-truth theory has been proposed as a solution to the Sorites paradox. However, one perennial attack to degree-of-truth theory is that its logic - fuzzy logic - is *non-classical*. Inspired by Gödel (1933), I attempt to better degree-of-truth theory by *classicalizing* it. That is, I attempt to give an interpretation of fuzzy logic within classical logic enriched by degree operators {○, ◔, ◑, ◕, ●} - "it is of no/low/moderate/high/full degree that …". Intuitively, degree-of-truth is classicalized as classical bivalent truth-value and a largely independent notion of degrees. A formal semantics of this enriched classical logic is presented, from which two semantic consequences are derived. The two semantic consequences are applied to analyse the *(in)validity* of the Sorites argument. There are two results: 1. the validity of the standard Sorites argument is reasserted, 2. a new argument for the *in*validity of the *degreed* version of the Sorites argument is presented.

References

Dietz, Richard and Moruzzi, Sebastiano (eds.). (2009). *Cuts and Clouds: Vagueness, Its Nature, and Its Logic*. Oxford University Press

Edgington, Dorothy. (1992). "Validity, Uncertainty and Vagueness". *Analysis*, 52(4): 193–204.

Edgington, Dorothy. (1997). "Vagueness by Degrees". in edited Rosanna Keefe and Peter Smith. (1997). *Vagueness: A Reader*. MIT Press. p. 294–316.

Fine, Kit. (1975). "Vagueness, Truth and Logic". *Synthese*, 30: p. 265–300.

Forbes, Graeme. (1983). "Thisness and Vagueness." *Synthese*, 54: p. 235–59.

Forbes, Graeme. (1985). *The Metaphysics of Modality*. Clarendon Press, Oxford.

Frege, Gottlob. (1897). *Logic*. in edited Michael Beaney. (1997). *The Frege Reader*. Blackwell Publishing. p. 227–250.

Goguen, J. A. (1969). "The Logic of Inexact Concept". *Synthese* 19: p. 325–73.

Gödel, Kurt. (1933). "Eine Interpretation des intuitionistischen Aussagenkalküls", *Ergebnisse eines mathematischen Kolloquiums*, 4: 39–40. translated as "An interpretation of the intuitionistic propositional calculus" and reprinted in edited Solomon Feferman. (1986). *Kurt Gödel Collected Works Volume I: Publications 1929–1936*. p. 300–303.

Hyde, Mominic. (2008). *Vagueness, Logic and Ontology*, (Ashgate new critical thinking in philosophy). Aldershot: Ashgate.

King, J. L. (1979). "Bivalence and the Sorites Paradox." *American Philosophical Quarterly* 16: p. 17–25.

Klement, Erich Peter, Mesiar, Radko, and Pap, Endre. (2000). *Triangular Norms* (Trends in Logic, Volume 8). Springer: Kluwer Academic Publisher.

Machina, Kenton F. (1976). "Truth, Belief, and Vagueness." *Journal of Philosophical Logic*, vol. 5, no. 1, 1976, p. 47–78.

McGee, Vann and McLaughlin, Brian (1995). "Distinctions Without a Difference." *Southern Journal of Philosophy*, 33 (supp.): p. 203–51.

Lakoff, G. (1973). "Hedges: a Study in Meaning Criteria and the Logic of Fuzzy Concepts". *Journal of Philosophical Logic* 2: p. 458–508.

Łukasiewicz, Jan and Tarski, Alfred. (1930). "Untersuchungen über den Aussagenkalkül". *Comptes Rendus des Séances de la Sociélé des Scierices et des Lettres des Varsovie Classe* III, 23, p. 30–50. translated by J. H. Woodger as "Investigations into the Sentential Calculus" in Tarski. (1956). *Logic, Semantics, Metamathematics*. Oxford University Press. p. 38–59.

Priest, Graham. (2008). *An Introduction to Non-Classical Logic: From If to Is*. 2nd edition. Cambridge University Press.

Sainsbury, R. M. (1986). "Degrees of Belief and Degrees of Truth." *Philosophical Papers*, 15: p. 97–106.

Smith, Nicholas. (2008). *Vagueness and Degrees of Truth*. Oxford University Press.

Tarski, Alfred. (1936). "O pojęciu wynikania logicznego". *Przegląd Filozoficzny*, 39: p. 58–68. translated as "On the Concept of Following Logically" by M. Stroińska and D. Hitchcock (2002). *History and Philosophy of Logic*, 23: 155–96.

Tarski, Alfred. (1936). "O pojęciu wynikania logicznego". *Przegląd Filozoficzny*, 39: 58–68. translated as "On the Concept of Following Logically" by M. Stroińska and D. Hitchcock (2002). *History and Philosophy of Logic*, 23: 155–96.

Williamson, Timothy. (1994). *Vagueness*. London and New York: Routledge.

Zadeh, L. A. (1975). "Fuzzy Logic and Approximate Reasoning." *Synthese* 30: p. 407–28.

Non-exclusive licence to reproduce thesis and make thesis public

I, Ching Hei, Yau ,
(*author's name*)

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

A Classical Degree-Theoretic Treatment of the Sorites Paradox ,
(*title of thesis*)

supervised by Tuomo Juhani Yli-Vakkuri, and Bryan Reed Stewart Frances .
(*supervisors's name*)

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

author's name: Ching Hei, Yau
dd/mm/yyyy: 13/05/2019