“Almost Identical with Itself”

A Search for a Logic of Fuzzy Identity

Lassi Saario
Master’s Thesis
Theoretical Philosophy
Faculty of Arts
University of Helsinki
November 2020

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty
Faculty of Arts

Tekijä – Författare – Author
Lassi Saario

Työn nimi – Arbetets titel – Title
“Almost Identical with Itself”: A Search for a Logic of Fuzzy Identity

Oppiaine – Läroämne – Subject
Theoretical Philosophy

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Master’s Thesis	November 2020	79

Tiivistelmä – Referat – Abstract

This thesis grows out of a fascination with the vagueness of natural language, its manifestation in the ancient Sorites paradox, and the way in which the paradox is dealt with in fuzzy logic. It is an attempt to resolve the tension between two versions of the paradox, and the related problem of whether identity can be fuzzy. If it can be fuzzy, then the most popular argument against vague objects is mistaken, which would be great news for those who hold that there can be vagueness in the world independently of our representation or knowledge of it.

The standard Sorites is made up of conditionals about an ordinary predicate (e.g. “heap”) by the rule of modus ponens. It is typically solved in fuzzy logic by interpreting the predicate as a fuzzy relation and showing that the argument fails as a result. There is another, less known version of the paradox, based on the identity predicate and the rule of substitutivity of identicals. The strong analogy between the two versions suggests that their solutions might be analogical as well, which would make identity just as vague as any relation. Yet the idea of vague identity has traditionally been rejected on both formal and philosophical grounds. Even Nicholas J. J. Smith, who is known for his positive attitude toward fuzzy relations in general, denies that identity could be fuzzy. The opposite position is taken by Graham Priest, who argues for a fuzzy interpretation of identity as a similarity relation. Following Priest, I aim to show that there is a perfectly sensible logic of fuzzy identity and that a fuzzy theoretician of vagueness therefore cannot rule out fuzzy identity on logical grounds alone.

I compare two fuzzy solutions to the identity Sorites: Priest’s solution, based on the notion of local validity, and B. Jack Copeland’s solution, based on the failure of contraction in sequent calculus. I provide a synthesis of the two solutions, suggesting that Priest’s local validity counts as a genuine kind of validity even if he might not think so himself. The substitutivity of identicals is not locally valid in Priest’s logic, however; his solution only applies to a special case with the rule of transitivity. Applying L. Valverde’s representation theorem and other mathematical results, I lay the foundation for a stronger logic where the substitutivity rule is locally valid and the two Sorites merge into one paradox with one solution. Finally, I defend fuzzy identity against Gareth Evans’ argument that vague identity leads to contradiction, and Smith’s argument that vague identity is not really identity. The former relies on a fallacious application of the substitutivity rule; to the latter, my principal response is to question Smith’s understanding of identity and argue for a broader one. I conclude that not only is fuzzy identity logically possible, but it also has potential applicability in metaphysics and elsewhere.

Avainsanat – Nyckelord – Keywords
Fuzzy Logic, Vagueness, Sorites Paradox, Identity, Similarity, Leibniz’s Law

Säilytyspaikka – Förvaringställe – Where deposited
City Centre Campus Library

Muita tietoja – Övriga uppgifter – Additional information

Contents

Introduction

Can there be vague objects? At least there are many objects in the world that *appear* more or less vague. By a vague object, I mean an object that has vague[1] spatial, temporal or modal (transworldly) boundaries because of the way it is in itself, independently of our representation or knowledge of it (cf. Paganini, 2011, p. 351). Putative examples of vague objects include clouds (Unger, 1980), mountains (Tye, 1990), cars (Forbes, 2010), tables (Kripke, 1972/1980, p. 51, n. 18), and cats (Lewis, 1993).

Vague objects are intimately tied with vague identity. By vague identity, I do not mean vague *statements* of identity, such as "Princeton = Princeton Borough" (Lewis, 1988), which is vague only because "Princeton" is an imprecise designator. Rather, I mean a relation of identity that is vague because of the related objects themselves. For instance, consider the identity of "Everest" and "Rock". Everest is the famous mountain, alleged to have vague boundaries; Rock is a hunk of rock, snow and ice, contained within crisp boundaries that do not determinately coincide with those of Everest. It can be argued that the identity "Everest = Rock" is vague; not because the terms "Everest" and "Rock" fail to refer to determinate objects in the world, but because the objects they refer to are only vaguely identical (Noonan & Curtis, 2018, § 9).

There is a popular argument against vague objects, based on the very connection between vague objects and vague identity. The best-known form of that argument is the one given by Evans (1978) and further explained by Lewis (1988). Evans argues that if there are vague objects, then there must be vague identity, but vague identity entails a contradiction; therefore, there cannot be vague objects (see Paganini, 2011, p. 352). It follows that the apparent vagueness of objects cannot be about the objects themselves; rather, it is about the words that denote them, or about our ignorance of their precise boundaries.

It is here that this thesis steps in. It focuses on Evans's second premiss, according to which the idea of vague identity is incoherent. It builds on the so-called fuzzy approach to vagueness, where vagueness is taken to be a matter of degree, explicated in fuzzy logic. On this approach,

---

[1] What exactly I mean by 'vague' will become clear by Section 2.1. The terms 'vagueness' and 'indeterminacy' are used somewhat confusingly in the literature, but it seems common to think of the latter as the more general phenomenon implicit in the former (see Barnes, 2010, p. 623, n. 20).

the question about vague identity turns into one about *fuzzy identity*. An identity relation $=$ is fuzzy if there are objects a, b such that $a = b$ is true only to an intermediate degree on the continuum from full falsity to full truth. If we accept that vague objects require fuzzy identity, the big question then becomes, can there be fuzzy identity?

Given the traditional resistance against vague identity, it comes as no surprise that there is resistance against fuzzy identity as well. Even Nicholas J. J. Smith, the perhaps most notable fuzzy theorist of vagueness in our times, rejects the view that identity could be fuzzy (Smith, 2008b). Identity is thought to be the relation in which every thing stands to itself and only itself, and nothing can be more precise than that. Yet in fuzzy logic, fuzzy relations astoundingly reminiscent of classical identity have long been studied under such labels as 'similarity relation,' 'fuzzy equivalence relation,' and 'indistinguishability operator.' Graham Priest (1998; 2008) has suggested that a relation of this kind would indeed be the identity of vague objects.

Both Smith and Priest take the fuzzy approach to vagueness, but their views conflict when it comes to the vagueness of identity. Whereas Priest treats identity as a fuzzy relation like any other, Smith considers it a special case that cannot be fuzzy even if all other relations can. Vague identity thus remains controversial even inside the fuzzy approach. The disagreement seems to stem from an intrinsic tension in the notion of identity, reflected in the oxymoronic title "Almost identical with itself." Identity seems possibly fuzzy in one sense, and necessarily crisp in another. Is there a way to combine the two points of view? Specifically, can the fuzzy approach to vagueness accommodate fuzzy identity, or is it actually the case that it *must* accommodate fuzzy identity if it is to accommodate fuzzy relations at all? It is this question that my thesis will aim to answer.

There is a related tension between the logical and the metaphysical approach to vagueness. Much of the criticism of vague identity is metaphysical of nature, focused against the idea that there could be vagueness in the world. It has been argued that the world as such is perfectly precise: we just fail to describe it precisely, or to have precise knowledge about it. Even if this view were justified on metaphysical grounds, I strongly feel that such considerations should not limit a logician's imagination. I think this ethos is nicely expressed by Sorensen (2018, § 8):

> In the absence of a decisive *reductio ad absurdum*, many logicians feel their role to be the liberal one of articulating the logical space for vague objects. There should be "Vague objects for those who want them" (Cowles and White 1991). Logic should be ontologically neutral.

Sorensen is here referring to Cowles and White's (1991) response to Pelletier's (1989) argument against vague objects. Pelletier's argument is that for there to be vague objects there ought to be a logic of a certain sort, but a logic of this sort entails a contradiction; therefore, there are no vague objects. Cowles and White express their suspicion toward drawing such a strong metaphysical conclusion from essentially logical premisses. They argue that there is a perfectly consistent logic which makes sense of vague objects. And yet their aim is not to argue that there *are* vague objects.

Likewise, my aim is to argue that there is a perfectly consistent logic which makes sense of fuzzy identity, and yet it is not to argue that there *is* fuzzy identity, let alone vague objects. Rather, it is to argue that there could be fuzzy identity *if* the proposed logic were taken seriously—whether it should be taken seriously is a separate question that will not be addressed here. The challenge posed by worldly vagueness will of course inspire the search for the logic, but to show that it has metaphysical applications is however secondary to the primary goal, which is to show that there are no *logical* obstacles for fuzzy identity. As for Evans's argument, that is to show that the second premiss fails and that the incoherence of vague identity is therefore no reason to rule out vague objects.

My approach to logic will alternate between a philosophical and a mathematical one: philosophical problems will motivate the formulation and derivation of mathematical notions and results, which will subsequently be subjected to philosophical evaluation. The reader is expected to know the basics of classical (first-order) logic, set theory, and algebra. Because of the metaphysical motivation, my emphasis will naturally be on the models of fuzzy identity and the mathematical theory of similarity relations employed in such models. Reasoning with fuzzy identity will be addressed in the context of the Sorites paradox, but the detailed study of proof-theoretical systems must be left for another occasion.

I will first introduce the formalisms that are needed throughout the thesis (Chapter 1). I will then introduce the general phenomenon of vagueness, the fuzzy approach to vagueness, and the controversy about vague identity (Chapter 2). I will survey the prominent approach to fuzzy identity as a similarity relation, focusing on Priest's theory of fuzzy identity (Chapter 3) and further developing it to a direction that makes it more comparable to classical identity (Chapter 4). Finally, I will weigh the final proposal against expectable objections and consider its philosophical implications (Chapter 5).

Chapter 1

Fuzzy logic

The roots of fuzzy logic are in the infinitely-valued logic of Lukasiewicz (Łukasiewicz & Tarski, 1930), the Gödel–Dummett logic (Gödel, 1932; Dummett, 1959), and the fuzzy set theory of Zadeh (1965). In the broad sense, fuzzy logic is a fruitful, growing area of research with a wide range of applications from computer science and engineering to social sciences and linguistics. Fuzzy logic in the narrow sense is a branch of mathematical logic, but even in this narrow sense 'fuzzy logic' is actually a family of logics, the common feature of which is replacing the classical truth values $\{0, 1\}$ with the real unit interval $[0, 1]$.

It is common to represent fuzzy logics as based on 't-norms,' which can be thought of as generalizations of the classical conjunction to $[0, 1]$.[1] What follows is a very brief introduction to a very small (but important) portion of fuzzy logics: logics of *continuous* t-norms. The exposition is mostly based on that of Běhounek et al. (2011), Priest (2008, Ch. 11, 25) and Smith (2008a, Ch. 1.2).

1.1 Algebras of continuous t-norms

Definition 1.1.1 (T-norm). A binary function $*\colon [0,1]^2 \to [0,1]$ is a *triangular norm*, or *t-norm* for short, if for all $x, y, z \in [0, 1]$,

- $x * y = y * x$ (*commutativity*)
- $(x * y) * z = x * (y * z)$ (*associativity*)
- $x \leq y$ implies $x * z \leq y * z$ (*monotonicity*)
- $x * 1 = x$ (*unit*)

---

[1] This is the so-called 'intended semantics' of fuzzy logics. They can also be studied with respect to more general algebraic semantics, with an abstract set of truth values.

In this thesis, we will restrict ourselves to *continuous* t-norms. For each continuous t-norm $*$, there is a unique binary function $\Rightarrow_*$ on $[0,1]$ satisfying the *residuation condition*

$$z * x \leq y \quad \text{iff} \quad z \leq x \Rightarrow_* y \tag{RC}$$

The function $\Rightarrow_*$ is called the *residuum* of $*$, and defined explicitly by

$$x \Rightarrow_* y = \max\{z \mid x * z \leq y\}$$

Each residuum in turn determines a *bi-residuum* and a *residuated negation*:

$$x \Leftrightarrow_* y = \min\{x \Rightarrow_* y, y \Rightarrow_* x\} \qquad\qquad \neg_* x = x \Rightarrow_* 0$$

In this way, every continuous t-norm $*$ defines a family of functions on the set of truth-values. These functions, together with minimum, maximum, and the constants 0 and 1, equip the set of truth values with a structure that is called the *t-algebra* of $*$:

$$[0,1]_* = ([0,1], *, \Rightarrow_*, \min, \max, 0, 1)$$

In the construction of the logic of $*$, the functions $*$, $\Rightarrow_*$, $\Leftrightarrow_*$, $\neg_*$, min and max will define truth conditions for the logical connectives & (the *strong* conjunction), $\rightarrow$, $\leftrightarrow$, $\neg$, $\wedge$ (the *weak* conjunction) and $\vee$, respectively.[2] (We shall occasionally use the symbols $\wedge$ and $\vee$ for the algebraic operations instead of min and max, for ease of exposition.)

There are three continuous t-norms of particular interest:

- The *Gödel t-norm*: $x *_G y = \min\{x, y\}$
- The *product t-norm*: $x *_\Pi y = x \cdot y$
- The *Łukasiewicz t-norm*: $x *_Ł y = \max\{x + y - 1, 0\}$

These three t-norms are the basic building blocks of fuzzy logic in the sense that every continuous t-norm can be represented as an *ordinal sum* of them.[3] The explicit formulae for their residuum, bi-residuum and residual negation are summarized in Table 1.1.

---

[2] As a matter of fact, it is sufficient to have $*$, $\Rightarrow_*$ and 0 as primitive functions and define the other functions and connectives by them (see Běhounek et al., 2011, p. 10).

[3] This result is known as the Mostert–Shields Theorem (Mostert & Shields, 1957).

	G	Π	Ł
$x \Rightarrow_* y$ (for $x > y$)	y	$\dfrac{y}{x}$	$1 - x + y$
$x \Leftrightarrow_* y$ (for $x \neq y$)	$\min\{x, y\}$	$\min\left\{\dfrac{x}{y}, \dfrac{y}{x}\right\}$	$1 - \lvert x - y \rvert$
$\neg_* x$	$\begin{cases} 1 & \text{for } x = 0 \\ 0 & \text{for } x > 0 \end{cases}$		$1 - x$

Table 1.1: The residuum, bi-residuum and residual negation of the t-norms $*_G$, $*_\Pi$ and $*_Ł$. Observe that all t-norms have $x \Rightarrow_* y = 1$ for $x \leq y$ and $x \Leftrightarrow_* y = 1$ for $x = y$.

1.2 Logics of continuous t-norms

A language $\mathcal{L}$ of (first-order) fuzzy logic contains the following logical symbols: the individual variables $x, y, z, \ldots$; the propositional constants $\top$ and $\bot$; the connectives $\neg$, &, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$; the quantifiers $\forall$, $\exists$; and the parentheses (,). In addition to the logical symbols, $\mathcal{L}$ may contain following non-logical symbols: individual constants $a, b, c, \ldots$; n-ary function symbols $f^n, g^n, h^n, \ldots$; and n-ary predicate symbols $P^n, Q^n, R^n, \ldots$. (Superscripts may be omitted when the arity is known from the context.)

The $\mathcal{L}$-terms and $\mathcal{L}$-formulae are defined inductively in the usual way. Variables are terms, constants are terms, and if $t_1, \ldots, t_n$ are terms, then $f^n(t_1, \ldots, t_n)$ is also a term. If $t_1, \ldots, t_n$ are terms, P^n a predicate symbol, A and B formulae, and x a variable, then $P^n(t_1, \ldots, t_n)$, $\neg A$, $(A \mathbin{\&} B)$, $(A \wedge B)$, $(A \vee B)$, $(A \rightarrow B)$, $(A \leftrightarrow B)$, $\forall x A$ and $\exists x A$ are formulae. (Parentheses may be omitted when this will cause no confusion.)

In the quantified formulae $\forall x A$ and $\exists x A$, the formula A is called the *scope* of the quantifier. An instance of a variable x is said to be *bound* if it occurs within the scope of a quantifier, or if it is the x between the quantifier and its scope; otherwise it is said to be *free*. We write $A(x)$ to indicate that the variable x may occur free in A, and $A(t)$ to indicate that the term t has been substituted for the free occurrences of x in A. Such a substitution requires that t *is free for* x *in* A, i.e., that no free variable in t gets bound as a result. A formula is *closed* if it contains no free variables; otherwise it is *open*. A closed formula is also called a *sentence*.

We will now define the standard semantics for the language $\mathcal{L}$. Let $*$ be a continuous t-norm and $[0,1]_*$ the t-algebra of $*$.

A *model* $\mathcal{M}$ of $\mathcal{L}$ over $[0,1]_*$, or a $*$-*model* of $\mathcal{L}$, is a pair (D, I), where D is a non-empty domain of objects, and I is an interpretation function that assigns an object $I(c) \in D$ to every individual constant c, a function $I(f^n) \colon D^n \rightarrow D$ to every function symbol f^n, and a relation

$I(P^n) \subseteq D^n$ to every predicate symbol P^n. A relation $I(P^n) = R$ on D is a *fuzzy set* of n-tuples $(a_1, \ldots, a_n) \in D^n$, identified with a characteristic function $f_R : D^n \longrightarrow [0, 1]$ which maps each n-tuple of elements of D to its degree of membership in the set R. If f_R maps all tuples to the classical truth values $\{0, 1\}$, we say that R is *crisp* (or *classical*).

A *variable assignment* v is a function which assigns an object $v(x) \in D$ to every individual variable x. Given an assignment v, we can define another assignment $v(a/x)$ which assigns $a \in D$ to x but is otherwise just like v.

Given a model $\mathcal{M}$ and an assignment v, every term t is assigned a *denotation* $t_v^{\mathcal{M}} \in D$, defined as usual: $x_v^{\mathcal{M}} = v(x)$ for variables, $c_v^{\mathcal{M}} = I(c)$ for constants, and $f(t_1, \ldots, t_n)_v^{\mathcal{M}} = I(f)\left({t_1}_v^{\mathcal{M}}, \ldots, {t_n}_v^{\mathcal{M}}\right)$ for functional expressions. (We will not always distinguish between a term and its denotation if this is irrelevant to what we are trying to show.)

Given a model $\mathcal{M}$ and an assignment v, every formula A is assigned a *truth value* $\|A\|_v^{\mathcal{M}} \in [0, 1]$. The classical truth values 0 and 1 may now be interpreted as full falsity and full truth, and the intermediate values as the infinitely many *degrees* of truth between the two. Truth values for atomic formulae are defined as

$$\|R(t_1, \ldots, t_n)\|_v^{\mathcal{M}} = I(R)({t_1}_v^{\mathcal{M}}, \ldots, {t_n}_v^{\mathcal{M}})$$

Truth values for compound formulae are

$$\begin{aligned}
\|\top\|_v^{\mathcal{M}} &= 1 & \|A \,\&\, B\|_v^{\mathcal{M}} &= \|A\|_v^{\mathcal{M}} * \|B\|_v^{\mathcal{M}} \\
\|\bot\|_v^{\mathcal{M}} &= 0 & \|A \rightarrow B\|_v^{\mathcal{M}} &= \|A\|_v^{\mathcal{M}} \Rightarrow_* \|B\|_v^{\mathcal{M}} \\
\|\neg A\|_v^{\mathcal{M}} &= \neg_* \|A\|_v^{\mathcal{M}} & \|A \leftrightarrow B\|_v^{\mathcal{M}} &= \|A\|_v^{\mathcal{M}} \Leftrightarrow_* \|B\|_v^{\mathcal{M}} \\
\|A \wedge B\|_v^{\mathcal{M}} &= \min\{\|A\|_v^{\mathcal{M}}, \|B\|_v^{\mathcal{M}}\} & \|\forall x A\|_v^{\mathcal{M}} &= \inf\{\|A\|_{v(a/x)}^{\mathcal{M}} \mid a \in D\} \\
\|A \vee B\|_v^{\mathcal{M}} &= \max\{\|A\|_v^{\mathcal{M}}, \|B\|_v^{\mathcal{M}}\} & \|\exists x A\|_v^{\mathcal{M}} &= \sup\{\|A\|_{v(a/x)}^{\mathcal{M}} \mid a \in D\}
\end{aligned}$$

If A is a sentence, its truth value in the model $\mathcal{M}$ does not depend on the assignment v, which allows us to write simply $\|A\|^{\mathcal{M}}$ instead of $\|A\|_v^{\mathcal{M}}$. If Γ is a set of sentences, the set of truth values of those sentences can be abbreviated as $\|\Gamma\|^{\mathcal{M}} = \{\|A\|^{\mathcal{M}} \mid A \in \Gamma\}$. A set of sentences in a language $\mathcal{L}$ is also called an $\mathcal{L}$*-theory*.

A formula $A(x)$ defines a subset of D with the characteristic function $f_{A(x)}^{\mathcal{M},v} : D \longrightarrow [0, 1]$,

$$f_{A(x)}^{\mathcal{M},v}(y) = \|A(x)\|_{v(y/x)}^{\mathcal{M}}$$

When formulae are interpreted as sets, the logical operations on formulae defined above reduce to set-theoretic operations on the respective fuzzy sets. (Sets in turn are identified with their characteristic functions, and they can also be interpreted as properties of the elements of D, which allows us to often speak of formulae, sets, functions and properties interchangeably.) The union and intersection of two sets correspond to disjunction and (weak) conjunction, and their generalizations for arbitrary families of sets to existential and universal quantification:

$$(f \cup g)(x) = \max\{f(x), g(x)\} \qquad \bigcup\{f_i\}(x) = \sup_i\{f_i(x)\}$$

$$(f \cap g)(x) = \min\{f(x), g(x)\} \qquad \bigcap\{f_i\}(x) = \inf_i\{f_i(x)\}$$

In order to complete the definition of the logic of $*$, we must yet define the corresponding consequence relation. This is most naturally done by taking an $\varepsilon \in [0, 1]$ to be the minimum truth value required of acceptable sentences, so that $[\varepsilon, 1] \subseteq [0, 1]$ is the set of *designated* truth values. A sentence A is *acceptable* (or simply *true*) in a model $\mathcal{M}$, written as $\mathcal{M} \vDash_\varepsilon A$, if $\|A\|^{\mathcal{M}} \geq \varepsilon$. A set of sentences Γ is acceptable in a model, $\mathcal{M} \vDash_\varepsilon \Gamma$, if $\mathcal{M} \vDash_\varepsilon A$ for each $A \in \Gamma$. Each choice of ε gives rise to an associated notion of logical consequence and validity:

Definition 1.2.1. A sentence A is a *logical consequence* of the set of sentences Γ with respect to $\varepsilon \in [0, 1]$, written as $\Gamma \vDash_\varepsilon A$, if $\mathcal{M} \vDash_\varepsilon \Gamma$ implies $\mathcal{M} \vDash_\varepsilon A$ for every $*$-model $\mathcal{M}$, i.e., if all $*$-models preserve the designated truth values from the premisses to the conclusion. If there is a derivability relation $\vdash$ from Γ to A, we say that the derivation $\Gamma \vdash A$ is *valid* if $\Gamma \vDash_\varepsilon A$.

If the derivation is not only valid but also all the premisses are acceptable on the intended model, we may say that it is *sound*. If a sentence A is a logical consequence of the empty set, i.e., if $\mathcal{M} \vDash_\varepsilon A$ for all $*$-models $\mathcal{M}$, we write $\vDash_\varepsilon A$ and say that A is a *logical truth* or *valid*. Finally, sentences A and B are said to be *logically equivalent*, $A \equiv B$, if they always have the same truth value, i.e., if $\|A\|^{\mathcal{M}} = \|B\|^{\mathcal{M}}$ for all $*$-models $\mathcal{M}$.

A stronger notion of logical consequence and validity is defined as follows:

Definition 1.2.2 (Global validity)**.** Let A be a sentence, Γ a set of sentences, and $\vdash$ a derivability relation. If $\Gamma \vDash_\varepsilon A$ for all $\varepsilon \in [0, 1]$, we write $\Gamma \vDash A$ and say that A is a global (logical) consequence of Γ, and that the derivation $\Gamma \vdash A$ is *globally valid*.[4]

Observe that $\Gamma \vDash A$ is equivalent to the condition that $\inf(\|\Gamma\|^{\mathcal{M}}) \leq \|A\|^{\mathcal{M}}$ for every $*$-model $\mathcal{M}$, which in turn is equivalent to $\vDash_1 (B_1 \wedge \cdots \wedge B_n) \rightarrow A$ provided that $\Gamma = \{B_1, \ldots, B_n\}$.[5]

Obviously, the resulting logic depends on the chosen consequence relation, which in turn depends on the underlying t-norm $*$. When restricted to the classical truth values $\{0, 1\}$,

---

[4] For the idea behind this term, see Priest, 1998, pp. 334–336, and Section 3.2.

[5] See Priest, 2008, pp. 226–227.

however, the truth conditions are equivalent with the classical ones regardless of the t-norm: the truth conditions of & are then equivalent with those of $\wedge$. Classical logic may be obtained by simply substituting $\{0, 1\}$ for $[0, 1]$ in the underlying algebra of any t-norm-based logic so that 1 becomes the only designated truth value.

Unless otherwise specified, we use the term 'fuzzy logic' to refer to the logic of all continuous t-norms, obtained through a generalization of the definition of logical consequence to the set of all such t-norms, with the level of acceptability set to $\varepsilon = 1$. In the literature on fuzzy logics, this logic is known as (Hájek's) *basic logic*. The logics of $*_G$, $*_\Pi$ and $*_Ł$ are called *Gödel logic*, *product logic* and *Łukasiewicz logic*, respectively. In subscripts and contexts where a t-norm specifies some object or property (as $*$ in $*$-transitivity and -similarity, to be defined in Section 3.1), we may simply write G (Gödel), Π (product) and Ł (Łukasiewicz) instead of $*_G$, $*_\Pi$ and $*_Ł$.[6]

1.3 Fuzzy logic with crisp identity

Let us add the binary predicate '=' to the logical symbols of our language, $\mathcal{L}$. Accordingly, let us add to the definitions of a formula and its truth value that $t_1 = t_2$ for any terms t_1, t_2 is also a formula, interpreted in a model $\mathcal{M}$ with an assignment v as

$$\|t_1 = t_2\|_v^{\mathcal{M}} = \begin{cases} 1, & \text{if } {t_1}_v^{\mathcal{M}} \text{ is } {t_2}_v^{\mathcal{M}} \\ 0, & \text{otherwise} \end{cases}$$

We have thus defined fuzzy logic with identity—or *equality*, as it is often called in mathematical contexts.[7] Compared to other predicates, identity enjoys a special status as a logical symbol, the interpretation of which is fixed in all models. On this interpretation, identity is a crisp relation: two terms t_1 and t_2 are identical iff they denote the same object in the domain. The relation remains unchanged if the set of truth values is restricted to $\{0, 1\}$. We may therefore call it not just a generalization of classical identity but indeed *the* classical identity, for it is the same old relation that has just been moved to a new setting.

Identity can be characterized as the "smallest" equivalence relation where each distinct object forms its own equivalence class. As an equivalence relation, identity has the properties

---

[6] Although there are axiomatic (and other) systems for these logics, with different kinds of soundness and (im)completeness with respect to the standard semantics just given—not to mention the general semantics studied in the literature—we will not go into details here. For an overview of proof theory for fuzzy logics, see Metcalfe et al., 2008.

[7] Observe that the first occurrence of '=' in the definition is a symbol of the object language, whereas the other one belongs to the metalanguage. The two notions must be kept apart.

of *reflexivity*, *symmetry* and *transitivity*, expressible in first-order logic as

$$x = x \tag{Ref}$$

$$x = y \to y = x \tag{Sym}$$

$$(x = y \mathbin{\&} y = z) \to x = z \tag{Tr\&}$$

The principles of the *substitutivity of identicals* and the *indiscernibility of identicals*, also known as *Leibniz's Law*,[8] can be expressed schematically as

$$(x = y \mathbin{\&} A(x)) \to A(y) \tag{SI\&}$$

$$x = y \to (A(x) \to A(y)) \tag{LL}$$

where A is any formula such that y is free for x in $A(x)$. All five principles are obviously valid in fuzzy logic with identity. They are valid for all continuous t-norms even when $\wedge$ is substituted for & in Tr& and SI&:

$$(x = y \wedge y = z) \to x = z \tag{Tr$\wedge$}$$

$$(x = y \wedge A(x)) \to A(y) \tag{SI$\wedge$}$$

In fuzzy logic *without* identity, it can be easily shown that Sym and Tr& (Tr∧) are logical consequences of Ref and SI& (SI∧). It can also be shown that SI& and LL are equivalent, and LL can be equivalently expressed as $x = y \to (A(x) \leftrightarrow A(y))$.[9] The corresponding rules of inference are globally valid.

The converse of LL, the *identity of indiscernibles*, is not even schematically expressible in first-order logic:

$$\text{If for all } A,\ A(x) \leftrightarrow A(y), \text{ then } x = y \tag{II}$$

Nor is it valid in first-order logic with identity, as it is perfectly possible to have a model where two non-identical objects satisfy all first-order formulae exactly to the same degree. II is expressible only in second-order logic (cf. below).

Although fuzzy logic does not change the semantics of identity from what it is in classical logic, it does slightly change the way we reason about it. Ref and SI& (or LL) are sufficient

---

[8] Sometimes this label is reserved for the converse principle, the *identity of indiscernibles* (see below), or for the conjunction of the two principles.

[9] To see this, consider the instance of LL where $A(z)$ is $B(z) \to B(x)$.

axioms for identity in systems for classical logic (Negri & von Plato, 2011, §§ 2.5, 6.3), but in fuzzy logic, a third axiom must be added to ensure its crispness (Běhounek et al., 2011, p. 87):[10]

$$x = y \vee \neg x = y \tag{Cr}$$

Neither these nor any other axioms are able to *define* identity in first-order logic, however. A theory may be complete in the sense that it contains all true sentences about identity, but no theory of identity is specific enough to have a unique model up to isomorphism (i.e., there is no *categorical* theory of identity in first-order logic). To see this, suppose that Γ is a theory of identity in a language $\mathcal{L}$, and $\mathcal{M} = (D, I)$ a model of Γ; pick then an arbitrary element $a \in D$, posit a new element $a' \notin D$, and define a new model $\mathcal{M}' = (D', I')$ where $D' = D \cup \{a'\}$ and I' is like I with the addition that $I'(s)(\dots, a', \dots) = I'(s)(\dots, a, \dots)$ for all function and predicate symbols $s \in \mathcal{L}$. We thus obtain an alternative model of Γ with a different interpretation of identity. While Γ is strong enough to define a crisp equivalence relation, $\sim$, such that for any model $\mathcal{M} = (D, I)$ of Γ and any assignment v from variables to D,

$$x_v^{\mathcal{M}} \sim y_v^{\mathcal{M}} \qquad \text{iff} \qquad \|A(x)\|_v^{\mathcal{M}} = \|A(y)\|_v^{\mathcal{M}} \text{ for all } A(x) \text{, where } y \text{ is free for } x,$$

the relation $\sim$ does not necessarily coincide with the real identity on D: it just equalizes elements that are logically *indistinguishable* in $\mathcal{L}$ (see Běhounek et al., 2011, p. 87).

The only thinkable way to uniquely define the 'intended' identity in first-order logic would require that we be able to express properties of "being identical with a" for each $a \in D$. If predicate symbols for such properties were added in the language, we could then substitute them for $A(x)$ in SI& or LL and so single out the intended interpretation of the identity predicate—but that would be circular of course, for the definiens would presuppose the definiendum. Identity is truly definable only in second-order logic, where quantification is allowed over arbitrary properties and not just the expressible ones (see Noonan & Curtis, 2018, § 2).

---

[10] It is redundant in Łukasiewicz logic, however.

Chapter 2

Vagueness, fuzziness, and identity

This chapter provides an extended introduction to three themes in contemporary analytical philosophy that are central to this thesis: vagueness, fuzziness, and identity. Although they can be and are often discussed independently of each other, it is their intersection that we are interested in. It will set the context in which our research question naturally arises, and it will also point out a natural way to answer it.

We will begin with a short introduction to vagueness through the ancient Sorites paradox, which is typically interpreted as a manifestation of the vagueness of a given predicate (Section 2.1). We will then show how fuzzy logic is typically applied to solve the Sorites paradox on the fuzzy approach to vagueness (Section 2.2). Finally, we will formulate a Sorites paradox for identity that is analogous to the first one, and discuss the possibility that its solution be analogous as well, which would make identity a fuzzy relation (Sections 2.3 and 2.4).

2.1 The Sorites paradox

Suppose you have piled up 10,000 grains of wheat. What you have is certainly a "heap." How many grains can you remove from it so that it remains a heap? If you remove one grain, you plausibly still have a heap. But if 9,999 grains make a heap, so do 9,998; and if 9,998 grains make a heap, so do 9,997. This kind of reasoning can be iterated over and over again until there is only one grain left. It seems wrong, however, to insist that a single grain would suffice to make a heap. At what point did the heap cease to exist?

This is one version of an ancient puzzle known as the *Sorites paradox* (from the Greek noun *soros* 'heap'), first presented by the Megarian philosopher Eubulides in the 4th century BCE.[1]

---

[1] Eubulides did also present another paradox of the same kind, known as the paradox of the Bald Man (*falakros*), which is essentially the same paradox as the one about a heap. The term 'Sorites paradox' has come to cover all paradoxes of this kind.

We will reformulate it in modern predicate logic in order to analyze its logical structure. Let us denote the predicate “is a heap” by H, the original heap with 10,000 grains by r_0, and the successive heaps by $r_1, r_2, \ldots, r_{9{,}999}$, so that for each n, r_n is (perhaps unintuitively) the compound of 10,000 – n grains.[2] The paradox now arises from the following argument:[3]

$$H(r_0), \forall n(H(r_n) \longrightarrow H(r_{n+1})) \vdash H(r_{9{,}999}) \tag{SS}$$

We will call the first and second premiss the minor and major premiss, respectively. The inference from the premisses to the conclusion can be laid out in natural deduction style as a series of steps by modus ponens:

$$\dfrac{\begin{array}{c}\dfrac{\dfrac{H(r_0)\qquad\dfrac{\forall n(H(r_n)\longrightarrow H(r_{n+1}))}{H(r_0)\longrightarrow H(r_1)}}{H(r_1)}\qquad\dfrac{\forall n(H(r_n)\longrightarrow H(r_{n+1}))}{H(r_1)\longrightarrow H(r_2)}}{H(r_2)}\\ \vdots\\ H(r_{9{,}998})\end{array}\qquad\dfrac{\forall n(H(r_n)\longrightarrow H(r_{n+1}))}{H(r_{9{,}998})\longrightarrow H(r_{9{,}999})}}{H(r_{9{,}999})}$$

The argument is classically valid. If the premisses are true, then the conclusion must also be true. Yet we are inclined to accept the premisses but deny the conclusion, which is a paradox. To get rid of the paradox we must either deny the minor premiss, deny the major premiss, deny that classical logic applies to this case, or accept the paradoxical conclusion.[4]

All our options seem problematic. The minor premiss seems undeniable, for a start: 10,000 grains certainly make a heap. It does not help to suggest that it would take more than 10,000 grains to make a heap, for the argument could have been started from a larger number just as well—unless one is ready to deny that anything ever is a heap. If we instead deny the major premiss, we are forced to admit that the removal of a grain can transform a heap into a non-heap, but it is very implausible that a single grain would make any difference. As for denying the applicability of classical logic, we must acknowledge that this kind of a paradox can be given for practically any predicate in natural language. Many philosophers still favor the assumption that natural language obey the principles of classical logic, and the argument is undoubtedly

---

[2] I acknowledge that it might be more natural to index the heaps with their number of grains from 10,000 downwards, and not with their position in the series from 0 upwards. The latter indexing is chosen to maintain formal compatibility with the other variants of Sorites to be discussed later.

[3] A careful reader will notice that the quantification over n in a constant symbol r_n is not strictly grammatical. The proper form would have a functional expression $r(n)$ instead. We will overlook the difference in order to simplify the notation (cf. Priest, 1991, pp. 293–294).

[4] Cf. Hyde & Raffman, 2018, § 3.

valid in the classical sense of the term. But then again, if we accept the premisses and the validity of the argument, we are forced to accept that a single grain makes a heap.

Although philosophers disagree on what is the correct solution to the paradox, they do typically agree that the paradox arises because the predicate "heap" is *vague* (Hyde & Raffman, 2018; Priest, 2008, pp. 221–222; Smith, 2008a, pp. 136, 167). Besides giving rise to Sorites paradoxes, vague predicates are typically characterized as predicates that have *borderline cases* (Sorensen, 2018; Smith, 2008a, p. 1). To say that something is a borderline case of a predicate is to say that it is unclear whether the predicate applies to it or not. A compound of 10 grains of wheat, for instance, could be considered a borderline case of a heap, as it is neither clearly a heap nor clearly a non-heap. Having borderline cases clearly distinguishes vague predicates from *precise* (or *crisp*) predicates, such as "prime number": for every number n, the proposition "n is a prime number" is univocally either true or false.

It is tempting to think that having borderline cases could be used not only to informally characterize vagueness but also to formally define what it is. The problem is that the notion of a borderline case also appears vague. Not only is it unclear whether a predicate applies to a borderline case, but it is also unclear whether it is unclear whether a predicate applies to a borderline case. Borderline cases have borderline cases, but borderline cases of borderline cases also have borderline cases and so on, *ad infinitum*. This phenomenon is called *higher-order vagueness*. If it is taken seriously, then the notion of vagueness is itself a vague notion, and the definition of vagueness depends on the theory of vagueness.[5]

2.2 The fuzzy approach to vagueness

One approach to vagueness is to question the principle of bivalence, according to which each statement is either true or false (but not both). While this principle holds in the area for which classical logic was originally designed (mathematics), the Sorites paradox can be taken as evidence that it does not apply to natural language. Many-valued logics challenge it by positing additional truth values between the classical ones, allowing a gradual transition from (full) truth to (full) falsity along the Sorites series. Going all the way with this idea, we get a logic with infinitely many degrees of truth—i.e., fuzzy logic.

Fuzzy logic has been applied to vagueness by e.g. Goguen (1968–1969), Lakoff (1973), Machina (1976), Priest (1998), and Smith (2008a).[6] Rather than relying on any particular theo-

---

[5] Prominent theories of vagueness include e.g. epistemicism (Williamson, 1994), contextualism (Graff, 2000), supervaluationism (Fine, 1975), subvaluationism (Hyde, 1997), and 'degree theories' based on many-valued logics such as fuzzy logic. For an overview, see e.g. Sorensen, 2018.

[6] See also Black, 1937, for an early precedent of the fuzzy approach.

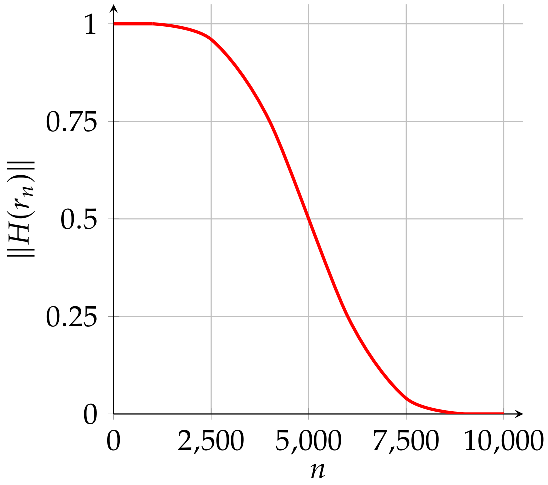

Figure 2.1: The truth value of $H(r_n)$ as a function of n.

retician or any particular theory of vagueness, we will here introduce two standard accounts of the Sorites paradox in variants of Łukasiewicz logic. The first account denies that the major premiss be true; the other one denies that the argument be valid (Williamson, 1994, pp. 123–124; Priest, 1998, p. 332; Smith, 2008a, pp. 265–266). The standard accounts are by no means perfect, their shortcomings being well-known and discussed in the literature, but they will anyway illustrate the nature and the appeal of the general framework within which this thesis is written.

As we recall, the Sorites paradox originates in classical logic. The available solutions in that framework are limited to denying either one of the premisses or accepting the conclusion, each of which is problematic in its own way. Once the framework is changed to fuzzy logic, the situation changes radically: it is then possible to deny the major premiss without accepting the counterintuitive claim that the removal of one grain can turn a heap into a non-heap.

In fuzzy logic, a conditional $H(r_n) \to H(r_{n+1})$ is fully true only if $H(r_n)$ is not truer than $H(r_{n+1})$; i.e., r_n is not more of a heap than r_{n+1} (recall that r_{n+1} is the compound with one grain less than r_n). The major premiss is fully true only if each conditional $H(r_n) \to H(r_{n+1})$ is fully true. The trick is to realize that not every such conditional is fully true. Now that we possess all the infinitely many degrees of truth between 0 and 1, there is no reason to stick to the classical intuition that the removal of a grain cannot make a compound any less a heap. Rather, it is plausible to assume that the removal of a grain can in fact make a difference as to whether the compound counts as a heap, never mind that it is a very slight difference. We may therefore suppose that there is some number m such that for every $n \leq m$, the truth value of $H(r_n)$ is exactly 1, but the truth value of $H(r_{m+1})$ is slightly less than 1 (say, 0.999), the truth value of $H(r_{m+2})$ still slightly less (say, 0.99), and so on, until the numbering reaches some k such that for every $n \geq k$, the truth value of $H(r_n)$ is exactly 0. Figure 2.1 illustrates the situation.

The precise diagnosis of the paradox depends on the preferred notion of validity. On the first account, validity is defined as the preserving of designated truth values (Definition 1.2.1), and full truth is recognized as the only designated truth value. As in classical logic, modus ponens does preserve full truth from the premisses to the conclusion: if $\|A\| = 1$ and $\|A \to B\| = 1$, then $\|B\| = 1$. The argument is thereby classified as valid, and yet it is unsound, because the major premiss is not fully true.

The question remains, however, why we were taken in by the paradox in the first place. What was so intuitive about the major premiss that we thought it was fully true although it was not? The answer is that while the major premiss is not fully true, *nor is it fully false either*. By the truth conditions of the Łukasiewicz conditional, $\|H(r_n) \to H(r_{n+1})\| = 1-(\|H(r_n)\|-\|H(r_{n+1})\|)$ whenever $\|H(r_n)\| > \|H(r_{n+1})\|$. If the maximum value of $\|H(r_n)\| - \|H(r_{n+1})\|$ is e.g. 0.01, it follows that the minimum value of $\|H(r_n) \to H(r_{n+1})\|$ is 0.99. Therefore, the major premiss is *almost* true, even though it is not fully true. We are so used to there being only two classical truth values that we fail to appreciate this slight difference. Perhaps we unconsciously "round up" the truth value from almost true to fully true, as it were.

The reliance on a premiss that is only almost true is one error, but the error also accumulates when it is repeated with every instantiation of the premiss. While the loss of truth may not be significant in a single step of inference, in a series of such steps some of the truth is lost in every step so that nothing might be left of it in the end. From the fully true and almost true premisses, we tend to draw the *fully false* conclusion, committing thus a fallacy of (almost) the worst kind.

On the other account, validity is taken to require the preserving of all truth values, i.e., global validity (Definition 1.2.2). As it turns out, modus ponens is not a globally valid rule of inference in Łukasiewicz logic. It does preserve full truth from the premisses to the conclusion, but if there are any premisses any less than fully true, it can happen that the conclusion is less true than the least true premiss. The argument is therefore invalid; as such, it is also unsound regardless of what truth values are designated. Why were we taken in, then, given that the argument is invalid? Perhaps we simply confused global validity with the alternative definition of validity as "full-truth-preserving," or mistakenly applied the classically valid modus ponens in a context where classical logic does not apply? More sophisticated explanations are developed in Sections 3.2 and 3.3.

2.3 Identity and the Sorites paradox

In Section 2.2, we saw a Sorites paradox for an ordinary (unary) relation and two alternative solutions based on fuzzy logic. We will now present a Sorites paradox for identity that will later (in Section 3.2) be solved in a way that likewise makes use of fuzzy logic. The paradox we present is a variant of the ancient paradox of Theseus's ship, first reported by Plutarch (1914, *Life of Theseus*, § 23). The paradox employs identity across time, but it could equally well be formulated across space, or possible worlds (Chisholm, 1967). Despite its superficial appearance, the paradox is not about any specific kind of identity but the identity of any real objects in general.

The ship is made up of 1,000 wooden planks that are replaced one by one until all the original planks have been replaced with new ones. Let us denote the original ship by s (= s_0) and the consecutive ships by s_n so that for each $n = 1, 2, \ldots, 1{,}000$, one of the original planks is replaced. These preconditions give rise to the following argument:

$$s = s_0, \forall n(s_n = s_{n+1}) \vdash s = s_{1,000} \tag{SD}$$

The minor premiss $s = s_0$ is self-evident, for it is logically true that an object is identical to itself.[7] The major premiss expresses the commonsensical position that the ship is still the same ship even if one of its planks is replaced. Together with the transitivity of identity, however, these premisses lead to the conclusion that the ship is still the same when all its planks have been replaced. The chain of transitivity from s_0 to $s_{1,000}$ can be analyzed as

$$\begin{array}{c}
\dfrac{\dfrac{s = s_0 \qquad \dfrac{\forall n(s_n = s_{n+1})}{s_0 = s_1}}{s = s_1} \qquad \dfrac{\forall n(s_n = s_{n+1})}{s_1 = s_2}}{s = s_2} \\
\vdots \\
\dfrac{s = s_{999} \qquad \dfrac{\forall n(s_n = s_{n+1})}{s_{999} = s_{1,000}}}{s = s_{1,000}}
\end{array}$$

There are of course many solutions available. One is to deny some of the identities $s_n = s_{n+1}$, but that would be to concede that there is an m such that $s = s_m$ but $s \neq s_{m+1}$. Whatever number is picked as m, the choice appears arbitrary—unless we deny *all* identities $s_n = s_{n+1}$, but that would be to deny the possibility of temporal identity and change altogether and hold each s_n as an independent entity. Another alternative is to accept the argument as such, which might be a natural solution to someone who takes formal continuity to be essential to identity. But even if this move may rescue one from the current paradox, it is not difficult to construct a slightly modified version that presents the same kind of problem but is not open to the same kind of solution. All one has to do is to imagine an alternative continuum where instead of

---

[7] From a logical point of view, this premiss is not even necessary for the argument, but it is maintained here to underline the analogy with the original Sorites paradox.

being replaced with a new one, each plank is simply removed and not replaced with anything. It still seems plausible that a ship can persist the removal of one plank, but after 999 steps there is only one plank left. At what point did the ship cease to exist?

The paradox displays interesting analogies with the standard Sorites paradox (SS). Both involve a series of objects $c_0, c_1, c_2, \dots,$ where the difference between any two objects c_n, c_{n+1} is negligible. Both make an allegedly self-evident claim about c_0 and suppose that it can be moved on along the series until it reaches an object for which it no longer holds, creating a paradox. In the first paradox, the "claim" is that c_n is a heap; in the second paradox, it is that c_n is identical with the original ship. The claim is transferred from each c_n to c_{n+1} by modus ponens in the first paradox, and by the transitivity of identity in the second one. The former rule operates on formulae, the latter on terms. The analogy becomes complete if we replace each $s_n = s_{n+1}$ with $s = s_n \rightarrow s = s_{n+1}$ in the second paradox: it then becomes a modus ponens inference of exactly the same kind as the first one, the only important difference being that "$s = \dots$" has taken the place of $H(\dots)$. Alternatively, we can strengthen the conditionals in the heap paradox to biconditionals with equal plausibility—just reason with the same Sorites series in a reverse order—to establish an analogy of a different kind. As Priest points out, "biconditionality is to sentences what identity is to terms, namely an equivalence relation satisfying substitutivity in extensional contexts" (1998, p. 331).[8]

The analogies can be taken to a more general level. Let us abstract away from the concrete arguments SS and SD to the respective schemata,

$$A(c_0), \forall n(A(c_n) \rightarrow A(c_{n+1})) \vdash A(c_m) \tag{SSS}$$

$$A(c_0), \forall n(c_n = c_{n+1}) \vdash A(c_m) \tag{SSD}$$

where $A(x)$ is a formula, $c_0, \dots, c_m$ are names for objects in a Sorites series, and $m \in \mathbb{N}$ is an arbitrarily large number. Both SSS and SSD can be analyzed as a chain of inferential steps as before. The former relies on modus ponens, the latter on the substitutivity of identicals. For each $i = 0, \dots, m-1$, the respective steps are of the forms

$$\dfrac{A(c_i) \qquad \dfrac{\forall n(A(c_n) \rightarrow A(c_{n+1}))}{A(c_i) \rightarrow A(c_{i+1})}}{A(c_{i+1})} \qquad\qquad \dfrac{A(c_i) \qquad \dfrac{\forall n(c_n = c_{n+1})}{c_i = c_{i+1}}}{A(c_{i+1})}$$

Now it is not hard to see how the two paradoxes are connected. If we take the bottom half

---

[8] For a detailed comparison of the two kinds of Sorites, see Priest, 1991; Priest, 1998, §§ 1, 4; Priest, 2008, §§ 25.5–25.6.

of the SSS step and the top half of the SSD step and put them together, we obtain

$$\dfrac{A(c_i) \qquad \dfrac{\dfrac{\forall n(c_n = c_{n+1})}{c_i = c_{i+1}}}{A(c_i) \to A(c_{i+1})}}{A(c_{i+1})}$$

where $A(c_i) \to A(c_{i+1})$ is derived from $c_i = c_{i+1}$ by Leibniz's Law. Chaining steps of this form together results in an alternative inference of SSD with the same core structure as SSS. (As for the other analogy, we could just as well derive by Leibniz's Law the biconditional $A(c_i) \leftrightarrow A(c_{i+1})$, which would match the strengthened form of the heap paradox.)

It is not even necessary to look at the actual steps of inference; we can just as well observe that the major premiss of SSS can be derived from that of SSD (or, semantically speaking, that the former is implied by the latter). Above, we already have a derivation from $\forall n(c_n = c_{n+1})$ to $A(c_i) \to A(c_{i+1})$ for an arbitrary i, whence we get $\forall n(A(c_n) \to A(c_{n+1}))$ by the introduction of the universal quantifier. Therefore, each paradox of the form SSS can be derived from SSD. Should we conclude that instead of two analogous kinds of paradoxes, we really have just one kind of paradox that comprises both?

Priest (1991) makes similar observations about the case where $A(x)$ is a monadic predicate. He argues that not only is the standard Sorites derivable from the identity Sorites, but that the converse also holds on "certain reasonable assumptions" (p. 294). As for their solutions, he writes:

> It is, of course, possible that the two kinds of Sorites have different kinds of solutions. However, the close structural similarity between the two kinds, makes this feel very implausible. Indeed, the fact that paradoxes of the two sorts are interderivable makes it very reasonable to suppose that the same sort of solution will apply to both kinds. (Priest, 1991, p. 295)

The way in which Priest converts a standard Sorites to an identity Sorites is far from straightforward, and the result is by no means as profound as the other way around. Be that as it may, I agree on his overall remarks about the strong similarity between the two kinds. Given that the identity Sorites is so similar to the standard one, it is tempting to think that its solution would be similar, too. If the illness is the same, so is the treatment.

Is it the same, however? Even if identity appears as vulnerable to Sorites paradoxes as any relation, we will see that its special position among relations makes its vagueness more difficult for many philosophers to accept. As similar as the two paradoxes may seem, the one with identity proves to be more fundamental and controversial in many ways.

2.4 Identity and worldly vagueness

Theories of vagueness differ in what they take to be the fundamental source of vagueness. There are three basic options (see e.g. Barnes, 2010, pp. 603–604), the first of which is to hold that vagueness is due to how the world is represented in e.g. our perception, language, or thought. For instance, there could be vagueness in how names represent objects, or how predicates represent relations. The second option is to locate vagueness out there in the world, in the objects and relations themselves. The third is to think of vagueness as an illusion that arises due to the limitedness of our epistemic capacity: on this view, names and predicates appear vague because we do not *know* their precise referents and extensions.

Where does the fuzzy approach stand on this matter? At least it does not, as such, have anything to do with the epistemic approach, although it may be compatible with it (see MacFarlane, 2010). The same can be said about representational vagueness, if we reason as Smith (2008a, pp. 70–71) does. In a fuzzy model $\mathcal{M} = (D, I)$ of a language $\mathcal{L}$, the function I univocally assigns an interpretation $I(s)$ to each non-logical symbol $s \in \mathcal{L}$. The vagueness is fully located inside the model, in the fuzzy sets assigned to the predicate symbols; and models, taken literally, are pictures of what the world is like. If the fuzzy approach is correct, vagueness is thereby located in the world (Smith, 2008a, pp. 70–71).[9]

The idea of worldly vagueness has traditionally been rejected as incoherent and unintelligible. If there is any vagueness at all (i.e., other than ignorance about precise boundaries, which can hardly be considered as genuine vagueness), it is all representational. This has been something of a consensus among philosophers (e.g. Russell, 1923; Dummett, 1975; Evans, 1978; Lewis, 1986; Sainsbury, 1995), although there have been signs of rebellion against the orthodox view over the recent decades (van Inwagen, 1990; Parsons & Woodruff, 1995; Edgington, 2000; Parsons, 2000; Barnes & Williams, 2011). The idea that identity could be vague, especially, has been met with considerable criticism. Even Smith, who is otherwise open-minded about vagueness in the world—in properties, relations, composition, existence, the world as a whole, even objects in a certain sense (Rosen & Smith, 2004; Smith, 2005)—nevertheless denies the possibility that identity be vague (Smith, 2008b).

What is so special about identity that it cannot be vague? The comparison of the two Sorites might provide some insight. Recall how the heap paradox is solved on the fuzzy approach by taking "is a heap" to be a fuzzy predicate. Contrast this with the possibility that the ship paradox could be solved similarly by taking "is identical to" to be a fuzzy predicate. The

---

[9] Whether the fuzzy approach really is correct in this respect is not something that I will address in this thesis (recall what I said on p. 7). It should also be emphasized that not everybody agrees with Smith on this matter (see e.g. Forbes, 1983, p. 245).

important difference is that in the heap paradox, while the objects r_n are taken to satisfy the predicate H to different degrees, their identity is not questioned or addressed in any way. Whether the heap is still the *same* heap after some grains have been removed is irrelevant to the paradox. (Indeed, it is not even necessary to suppose that the objects $r_0, \dots, r_{9,999}$ succeed each other over time: they could just as well be spatially distinct objects existing at the same time.) This is not the case in the ship paradox, where the issue is precisely whether the ship is still the same after some planks have been replaced. The heap paradox is about properties of objects, whereas the ship paradox is about objects themselves.

The connection between identity and objects has famously been underlined in the slogan "no entity without identity" by Quine (e.g. 1969, p. 23). It is meant to express the principle that in order for something to exist, it must have precise identity conditions. From the vagueness of the identity conditions for every-day objects such as ships, Quine concludes that such objects do not really exist: they are nothing more than convenient fictions which help us to make sense of the world as we experience it but shatter to pieces as soon as they are subjected to serious metaphysical examination (as cited in Juti, 2001, pp. 138–139, 232–233).

This line of argument is at its strongest when vagueness is taken to be something so messy that it cannot be made sense of by means of formal logic. If we cannot properly define identity conditions for clouds, for instance, we can plausibly conclude that clouds do not really exist. However, to require that identity conditions be formally defined does not necessarily mean that they should be defined in classical logic. If we could come up with well-defined conditions for vague identity in a non-classical logic, that would certainly cast a shadow of doubt on the Quinean argument. Is this not exactly what could be done in fuzzy logic? Would it not be great news if we were not forced to give up all ordinary objects, after all?

To define identity for vague objects is a challenge, no doubt, but it is a challenge that we are willing to take on.

Chapter 3

Similarity as fuzzy identity

We have seen (in Section 1.3) that it is possible to maintain the classical definition of identity as a crisp relation within the framework of fuzzy logic. Now we turn to the question of whether classical identity can be generalized to a fuzzy relation that allows intermediate degrees of truth and obeys fuzzy versions of the classical laws of identity. A necessary condition for such a generalization is of course that it includes classical identity as a special case, i.e., reduces back to classical identity when restricted to $\{0, 1\}$.

I will begin this chapter with a brief introduction to the theory of similarity relations and its relationship with the theory of metric spaces (Section 3.1). I will then focus on Graham Priest's theory of fuzzy identity as a similarity relation and his analysis of the identity Sorites (Section 3.2), contrasting it with an alternative analysis by B. Jack Copeland (Section 3.3). Finally, I will critically discuss Priest's theory from within the fuzzy framework (Section 3.4), pointing out several shortcomings that I will then try to overcome in Chapter 4.

3.1 Similarity relations

Classical identity is the relation of "sameness," for $x = y$ just means that x denotes the *same* object as y. It can also be characterized as the equivalence relation where each object has its own equivalence class. A natural approach to generalize classical identity, then, is to generalize the notion of an equivalence relation and to think of such a generalization as "similarity," as opposed to sameness or "full similarity."

Definition 3.1.1 (Similarity relation)**.** A fuzzy relation $S \subseteq D^2$ with the characteristic function $f_S \colon D^2 \longrightarrow [0, 1]$ is a $*$-*similarity relation* for a t-norm $*$ if for all $x, y, z \in D$,

1. $f_S(x, x) = 1$ (*reflexivity*)

2. $f_S(x, y) = f_S(y, x)$ *(symmetry)*

3. $f_S(x, y) * f_S(y, z) \leq f_S(x, z)$ *($*$-transitivity)*

If the t-norm $*$ is known from the context or otherwise irrelevant, we may simply say that S is a *similarity relation*.

Given a similarity relation S on D, we obtain *similarity classes* analogous to the equivalence classes of a classical equivalence relation (Zadeh, 1971, pp. 188–189). The similarity class of $x \in D$, $S[x]$, is defined as

$$f_{S[x]}(y) = f_S(x, y)$$

for all $y \in D$. It can be interpreted as the fuzzy set of elements that are similar to x. If S is $\wedge$-transitive, it also induces a fuzzy partition on D that can be represented in the form of a partition tree, analogous to the quotient of a classical equivalence relation (Zadeh, 1971, pp. 187–188).

The general notion of $*$-transitivity comprises a family of different transitivities, one for each t-norm $*$. The relative ordering of t-norms determines the ordering of the respective transitivities. As the Gödel t-norm is the pointwise largest t-norm (i.e., $x * y \leq x *_G y$ for all $*$), G-transitivity implies $*$-transitivity for each t-norm $*$. The product t-norm is pointwise smaller than the Gödel t-norm but pointwise larger than the Łukasiewicz t-norm. The corresponding similarity relations are ordered accordingly. Even though the pointwise ordering of t-norms is only partial, it is sufficient to determine a rough hierarchy of similarity where each particular similarity relation occupies a place. Combined with the observation that classical equivalence relation is a special case of G-similarity, we obtain a classification of similarity relations with increasing degree of generality (see Figure 3.1 for illustration).

As before, we are particularly interested in the three prominent t-norms $*_G$, $*_\Pi$ and $*_Ł$. Their respective similarity relations are well-known in the literature, only with different names than the ones preferred here. Gödel similarity is known simply as a 'similarity relation' in Zadeh (1971); product similarity is what Menger (1951) calls a 'probabilistic relation'; and Łukasiewicz similarity is the 'likeness relation' of Ruspini (1982). Our notion of 'similarity relation,' used as a cover term for fuzzy equivalence relations for all t-norms as opposed to Zadeh's narrow understanding of the term, corresponds e.g. to the 'fuzzy equivalence relation' in Klawonn and Castro (1995). It also corresponds, with some reservations, to the 'indistinguishability operator' in Trillas and Valverde (1984) and the 'similarity function' in Priest (2008, p. 580).

The notion of a similarity relation is closely related to that of a distance metric, which is studied within the mathematical theory of metric spaces.

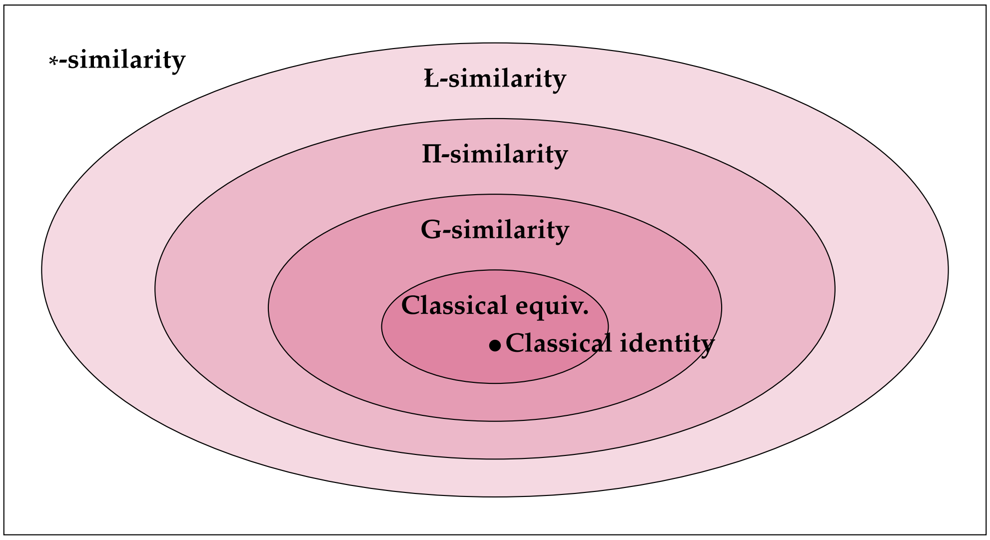

Figure 3.1: The hierarchy of (prominent) similarity relations.

Definition 3.1.2 (Distance metric)**.** A binary function $d\colon D^2 \to [0,1]$ is a *distance metric*[1] on the domain D if for all $x, y, z \in D$,

1. $d(x,x) = 0$
2. $d(x,y) = d(y,x)$
3. $d(x,z) \leq d(x,y) + d(y,z)$ (*triangle inequality*)

Oftentimes a fourth condition, $d(x,y) = 0 \Rightarrow x = y$, is added to ensure the non-degenerateness of the metric, but we can do without it for now (cf. Priest, 1998, n. 5; cf. also p. 58).

Intuitively, it is plausible that the similarity and the distance (understood in a broad, abstract sense) between two objects are in a converse relation to each other: when one goes up, the other goes down. The complement of a similarity relation—its corresponding 'dissimilarity' relation—is directly comparable to distance. It can be proved that the notions of a distance metric and Łukasiewicz dissimilarity are in fact equivalent. (In what follows, we will no longer use a separate symbol for the characteristic function of S so as to simplify the notation.)

Theorem 3.1.3. *Let d be a binary function from the domain D to $[0,1]$, and S a binary relation on D, such that $S(x,y) = 1 - d(x,y)$. Then d is a distance metric if and only if S is a Ł-similarity relation.*

Proof. If d is a distance metric, then the reflexivity and symmetry of S follow immediately from those of d, and vice versa. The triangle inequality of d and the Ł-transitivity of S are proved

---

[1] The scale of the metric in the definition is fixed as $[0,1]$. The scale could as well be defined as e.g. the set of non-negative real numbers, which is what Priest (1998) does, but we opt for a normalized metric as in Priest (2008, § 25.6) for the sake of simplicity.

equivalent as follows:

$$\begin{aligned} d(x,z) \leq d(x,y) + d(y,z) && \iff \\ 1 - S(x,z) \leq 1 - S(x,y) + 1 - S(y,z) && \iff \\ S(x,y) + S(y,z) - 1 \leq S(x,z) && \iff \\ \max\{0, S(x,y) + S(y,z) - 1\} \leq S(x,z) && \iff \\ S(x,y) *_{\text{Ł}} S(y,z) \leq S(x,z) && \square \end{aligned}$$

The mutual duality of a Łukasiewicz similarity relation and a distance metric can be generalized for a $*$-similarity and a $\oplus$-pseudometric, where $*$ is a continuous t-norm and $\oplus$ its dual *t-conorm* (Valverde, 1985, Th. 5.2). In particular, a Gödel similarity relation is dual to an *ultrametric*, where the condition of triangle inequality has been strengthened into ultrametric inequality (Zadeh, 1971, pp. 185–186).

3.2 Priest's theory of fuzzy identity

Graham Priest has proposed a theory of fuzzy identity as a solution to the Sorites paradox about identity. The theory is first presented in his article from 1998 and again in a slightly different form in his introductory book from 2008 (§ 25.6), where it is further considered as a theory of vague objects (§ 25.7).[2] Using the terminology of the previous section, we may roughly summarize Priest's theory as that of fuzzy identity as a Łukasiewicz similarity relation.

Let D be a domain of objects, and d a distance metric on D. We redefine a model as a triple $\mathcal{M} = (D, d, I)$, where I is an interpretation function. Terms and formulae are interpreted as before except for identity, reinterpreted as follows:

$$\|t_1 = t_2\|_v^{\mathcal{M}} = 1 - d\left({t_1}_v^{\mathcal{M}}, {t_2}_v^{\mathcal{M}}\right)$$

Identity is thereby interpreted as the complement of distance. If the distance between t_1 and t_2 is 0, then they are fully identical; if it is 1, then they are fully distinct (non-identical). The more they are distant, the less they are identical.

Our definition is a hybrid of the two slightly different definitions given in Priest (1998) and Priest (2008). One one hand, we presuppose with Priest (2008) that d is a normalized metric, i.e., that its codomain is $[0,1]$. Priest's original definition of d has the codomain $\mathbb{R}$, and the value of d is consequently bound in the definition of identity, but the difference is inessential

---

[2] See also Priest, 2003.

(see Priest, 1998, p. 333, n. 6). On the other hand, we maintain Priest's original conditions for a distance metric, the first of which has been strengthened in his later definition. This difference will become an object of our attention as we proceed (see p. 58).

From the definition of fuzzy identity, it follows immediately by Theorem 3.1.3 that fuzzy identity is a Łukasiewicz similarity relation. It therefore validates Ref, Sym and Tr& in Łukasiewicz logic. It does not validate Tr∧, however. To see this, suppose $d(a, b) = 0.3$, $d(b, c) = 0.4$ and $d(a, c) = 0.5$. The conditions of a distance metric are met; specifically, $d(a, c) \leq d(a, b) + d(b, c)$. It follows that $\|a = b\| = 0.7$, $\|b = c\| = 0.6$ and $\|a = c\| = 0.5$. Now $\|a = b\| \wedge \|b = c\| > \|a = c\|$, which contradicts the assumption that Tr∧ be valid.

Fuzzy identity also does not validate SI∧, SI& nor LL, which is equivalent to SI&. This is not surprising, given that the definition of fuzzy identity establishes no connection whatsoever between the distance metric d and the interpretation function I. If e.g. $d(x, y) = 0$, $f_P(x) = 1$ and $f_P(y) = 0$ for some $x, y \in D$ and $f_P = I(P)$, then $\|(x = y \mathbin{\&} P(x)) \rightarrow P(y)\| = 0$ regardless of the interpretation of &.

Priest (2008, pp. 575–576) acknowledges the failure of SI and goes on to propose a restriction on the predicate P, based on the fact that vague predicates, unlike crisp ones, are *tolerant*: a small change in the argument z causes only a small change in the truth value of $P(z)$. Priest defines that a predicate P is *smooth* if

$$|f_P(x) - f_P(y)| \leq d(x, y)$$

i.e., the change in the argument (as measured on the distance metric) sets an upper bound for the change in the applicability of the predicate. Suppose now that P is a smooth predicate. If $f_P(x) \geq f_P(y)$, then $|f_P(x) - f_P(y)| = f_P(x) - f_P(y) \leq d(x, y)$, and if $f_P(x) < f_P(y)$, then $f_P(x) - f_P(y) < 0 \leq d(x, y)$. Either way, we get

$$\begin{aligned}
f_P(x) - d(x, y) \leq f_P(y) &\qquad\qquad \Longleftrightarrow \\
\max\{f_P(x) - d(x, y), 0\} \leq f_P(y) &\qquad\qquad \Longleftrightarrow \\
\max\{1 - d(x, y) + f_P(x) - 1, 0\} \leq f_P(y) &\qquad\qquad \Longleftrightarrow \\
\|x = y\| *_{Ł} \|P(x)\| \leq \|P(y)\| &\qquad\qquad \Longleftrightarrow \\
\|(x = y \mathbin{\&} P(x)) \rightarrow P(y)\| = 1 &
\end{aligned}$$

We conclude that SI& holds for smooth predicates (or smooth formulae, if the definition of smoothness is generalized in the obvious way) in Łukasiewicz logic. SI∧ fails even when it is restricted to smooth predicates, but the corresponding rule of inference, as well as that of Tr∧, can yet be shown to be *locally valid* (see Definition 3.2.1). LL is valid for smooth predicates, and the corresponding rule $x = y \vdash P(x) \rightarrow P(y)$ is globally valid provided that P is smooth. However, if one adds the assumption $P(x)$ to conclude $P(y)$ by modus ponens, the overall inference fails to the same effect as SI.

We will now apply Priest's theory to the Sorites paradox about Theseus's ship (Section 2.3). For the sake of simplicity, we shall interpret the terms s_n directly as elements of the domain. The specific distance metric to be applied in our solution is

$$d(s_n, s_m) = \frac{|n - m|}{1{,}000}$$

It is easy to confirm that the metric thus defined meets the conditions of Definition 3.1.2. Intuitively, the distance between s_n and s_m can be thought of as the relative difference in their material composition or their position on the series of successive ships.[3] We are now able to calculate the truth values of the premisses and conclusion of the Sorites argument:

$$\begin{aligned} \|s = s_0\| &= 1 - d(s_0, s_0) = 1 - 0 = 1 \\ \|\forall n(s_n = s_{n+1})\| &= \inf_n \{1 - d(s_n, s_{n+1})\} = 1 - 0.001 = 0.999 \\ \|s = s_{1{,}000}\| &= 1 - d(s_0, s_{1{,}000}) = 1 - 1 = 0 \end{aligned}$$

We have thus derived a fully false conclusion from a fully true and an almost true premiss. The rule of transitivity of identity, just like modus ponens in the original paradox, is valid only in the "full-truth-preserving" sense: it preserves the level of acceptability ε only if $\varepsilon = 1$. The pattern is the same as in the original Sorites paradox, and so are its explanations (recall Section 2.2 and the comparison on p. 22). Depending on our notion of validity, we can say that the argument is either unsound (because the premisses are not acceptable) or invalid (because it is valid for $\varepsilon = 1$ only). We are taken in by the paradox because we are so blinded by classical logic that we mistakenly regard the almost true identities as fully true or the invalid rule as valid.

It can be argued that the invalidity explanation is not satisfactory, however. It was not so for the original Sorites, and it is not so for the present one either. Edgington (1992) points out that to classify modus ponens as invalid "obscures an important distinction: between the case where the fall in value of the conclusion is constrained by the values of the premisses, and the case where it is not" (p. 200). Copeland (1997) agrees:

---

[3] It is of course possible to define the distance in other ways as well. More plausible truth conditions might be something like the ones given for the heap paradox (recall Figure 2.1). However, the point here is not to argue for a specific truth definition of identity but for the more general idea that we can allow truth to have degrees and so solve the Sorites paradox.

> The all-or-nothing distinction between the logically perfect and the logically worthless that motivates the classical account of argument validity is completely foreign to fuzzy logic. . . . In general, where $A \rightarrow B$ is nearly true, B's value, although less than A's, will always be nearly the same as A's. Modus ponens is truth preserving enough. (Copeland, 1997, pp. 522–523)

The same point applies equally well to the transitivity of identity. Priest (1998) gives an exact formulation for what it is for an argument to be "truth preserving enough":

Definition 3.2.1 (Local validity)**.** An inference from a set of premisses Γ to a conclusion C is *locally valid* iff for every $\varepsilon \in [0,1)$, there is a $\delta \in [0,1)$ such that

$$\inf(\|\Gamma\|^{\mathcal{M}}) > \delta \quad \text{implies} \quad \|C\|^{\mathcal{M}} > \varepsilon$$

for all models $\mathcal{M}$ (Priest, 1998, pp. 334–335).

The idea is that if we have set the level of acceptability required of the conclusion, there is a level of acceptability required of the premisses that will guarantee the acceptability of the conclusion. In other words, the conclusion can be made arbitrarily close to full truth by bringing the premisses sufficiently close to full truth: "if our premisses are 'true enough,' the conclusion is going to be acceptable" (Priest, 1998, p. 335).

Every inference that is globally valid (Definition 1.2.2) is also locally valid, but the converse is not true. Modus ponens and the transitivity of identity are globally invalid but locally valid (Priest, 1998, pp. 334–335; Priest, 2008, pp. 575–576). The substitutivity of identity is locally valid for smooth predicates (Priest, 1998, pp. 337–338; Priest, 2008, pp. 575–576). It can also be shown that a chain of locally valid inferences is locally valid, which is a crucial result in explaining the nature of soritical reasoning (Priest, 1998, pp. 335–336).

The identity Sorites is a chain of inferential steps by the transitivity of identity. Therefore, it is locally valid. Let us see in detail why this is so. Suppose $\varepsilon \in [0,1)$ is the base level of acceptability. We start from the first step,

$$\dfrac{s = s_0 \qquad \dfrac{M}{s_0 = s_1}}{s = s_1}$$

where M is short for the major premiss $\forall n(s_n = s_{n+1})$. By the truth conditions of the formulae in question, it follows that the truth of the conclusion $s = s_1$ is bounded by the truths of the premisses $s = s_0$ and M,

$$\|s = s_0\| *_{\text{Ł}} \|M\| \leq \|s = s_0\| *_{\text{Ł}} \|s_0 = s_1\| \leq \|s = s_1\|$$

which effectively shows that the first step is locally valid. To formally prove this with respect to Definition 3.2.1, choose $\delta_1 = \frac{\varepsilon+1}{2}$. Now, if $\|s = s_0\|, \|M\| > \delta_1$, then

$$\varepsilon = \frac{\varepsilon+1}{2} + \frac{\varepsilon+1}{2} - 1 < \|s = s_0\| + \|M\| - 1 \leq \|s = s_0\| *_Ł \|M\|$$

and therefore $\|s = s_1\| > \varepsilon$.

We reason similarly with the second step to conclude $\|s = s_1\| *_Ł \|M\| \leq \|s = s_2\|$. Chaining the two steps together, we arrive at

$$\|s = s_0\| *_Ł \|M\|^2 \leq \|s = s_2\|$$

where $\|M\|^2 = \|M\| *_Ł \|M\|$. The two locally valid inferences are thus composed into an inference that is likewise locally valid, only the level of acceptability required of the premisses rises to $\delta_2 = \frac{\varepsilon+2}{3}$. If $\|s = s_0\|, \|M\| > \delta_2$, then

$$\varepsilon = \frac{\varepsilon+2}{3} + \frac{\varepsilon+2}{3} - 1 + \frac{\varepsilon+2}{3} - 1 < \|s = s_0\| + \|M\| - 1 + \|M\| - 1 \leq \|s = s_0\| *_Ł \|M\| *_Ł \|M\|$$

The same procedure is iterated to extend the chain with the third step, and so on. For each $n = 1, \ldots, 1{,}000$, we can show that the chain up to the nth step is locally valid with $\delta_n = \frac{\varepsilon+n}{n+1}$. Finally, we end up with

$$\|s = s_0\| *_Ł \|M\|^{1{,}000} \leq \|s = s_{1{,}000}\| \tag{3.1}$$

which amounts to the fact that the entire chain is locally valid. The level of acceptability required of the premisses is as high as $\delta_{1{,}000} = \frac{\varepsilon+1{,}000}{1{,}001}$.

The power $\|M\|^{1{,}000}$ plays a central role here. Unlike the classical conjunction, the Łukasiewicz t-norm is not idempotent: $x *_Ł x = x$ only if $x \in \{0, 1\}$, otherwise $x *_Ł x < x$. As the exponent n increases, the value of the power x^n decreases. This has the rather counterintuitive effect that repeating an assumption does in fact make it stronger. Assume M once, and you commit yourself to it only once; assume it again and it binds you twice as much.

Contrast 3.1 with the corresponding condition for global validity,

$$\|s = s_0\| \wedge \|M\|^{1{,}000} \leq \|s = s_{1{,}000}\|$$

where $\|M\|^{1{,}000} = \|M\| \wedge \cdots \wedge \|M\|$. The only difference with respect to local validity is having the minimum (Gödel t-norm) $\wedge$ in place of $*_Ł$. But the minimum, like the classical conjunction and unlike $*_Ł$, is idempotent—indeed, it is the only such t-norm—so we could just as well require $\|s = s_0\| \wedge \|M\| \leq \|s = s_{1,000}\|$. As natural as this move may appear, it is not available in Łukasiewicz logic. Perhaps this offers a new explanation for why we were taken in by the paradox? We mistakenly supposed that the conjunction in question was the classical idempotent one, overlooking its inapplicability to vague natural language. We confused it with the fuzzy Łukasiewicz conjunction because the difference in their truth values is negligible for small exponents, but as the exponent increases, so does the difference. While the classical conjunction remains at 0.999, the fuzzy one reduces by 0.001 every time the exponent is incremented by one and finally reduces to full falsity. A small error accumulates to a big one.

The above considerations on the transitivity chain of length 1,000 could be generalized for any Sorites chain of length K. We have already remarked (on p. 20) that the longer the chain gets, the more of the truth of the premisses is lost in the inference. We are now in a position to restate this point with added rigor. The higher the level of acceptability required of the conclusion, the shorter the chain and/or the truer the premisses will have to be. In the case of the identity paradox, the inference could be made acceptable for some ε if the value of K were made low enough or the value of δ high enough, but no value of ε (other than 0) is low enough to allow for K as high as 1,000 and δ as low as 0.999.

Once we have fixed the premisses (and their interpretation), the variable δ no longer plays any role in that particular argument; it is then K alone that sheds light on why the paradox arises. As Priest (1998, p. 336) puts it, "sorites inferences are acceptable provided we use them locally (to take us a short distance down the sorites), but not globally." This provides us with a new, more sophisticated explanation for why we were taken in by the paradox: "Our mistake is simply thinking that inferences that can be used over short distances are reliable over long distances. That is, we confuse (global) validity with local validity" (Priest, 1998, p. 336).

We must also be careful not to confuse local validity with soundness (recall p. 12). As in classical logic, validity of an argument only guarantees that the conclusion is acceptable *if* the premisses are—it does not guarantee that they are. In dealing with local validity, we may require different levels of acceptability from the premisses and the conclusion, but the basic distinction between validity and soundness remains the same. While the identity paradox relies on a form of inference that is locally valid in general, it is by no means sound in that particular case.[4]

---

[4] "That an inference is locally valid does not, . . . on its own, answer the question of whether it is applicable on a particular occasion" (Priest, 1998, p. 335).

3.3 Copeland's syntactic analysis

Priest (1998) writes in an endnote: "It is an interesting project to give a proof theory for the notion of local validity, but one for another occasion" (p. 341, n. 12). Ironically enough, at about the same time Copeland (1997) writes on the same topic, but unlike Priest, he provides a proof-theoretic analysis of the situation, whereas the semantics "must await another occasion" (p. 523). Luckily, we are in a position to combine Priest's semantic and Copeland's syntactic considerations to obtain a more complete understanding of the Sorites phenomenon.

(To be exact, Copeland considers the standard Sorites with modus ponens and not the identity version with transitivity, even if his topic is vague identity in fuzzy logic, and even if he takes it for granted that identity can be fuzzy. Given the analogy between the two paradoxes, we can translate his insights from the former to the latter in a straightforward manner.)

Copeland's analysis of the Sorites argument employs Gentzen's (1934/1969) sequent calculus, and his rule of cut in particular. The rules in Gentzen's calculus are inferences not between formulae but between *sequents* of the form $\Gamma \vdash \Delta$, where Γ and Δ are sequences (lists) of formulae. Unlike Gentzen, we will treat them as finite multisets, i.e., lists with multiplicity but no order, allowing us to avoid exchange rules (Negri & von Plato, 2001, pp. 14–15).

Sequents of the form $\Gamma \vdash C$, with only one formula on the right-hand side, can be given an *operational interpretation*: they can be taken to express that the conclusion C is derivable from the assumptions Γ (Negri & von Plato, 2001, p. 47). In a similar vein, the rule of cut,

$$\frac{\Gamma \vdash \Delta, A \qquad A, \Gamma' \vdash \Delta'}{\Gamma, \Gamma' \vdash \Delta, \Delta'}\ \mathit{Cut}$$

can be thought of as composing two derivations into one, provided that Δ is empty. This makes sequent calculus a useful tool for the analysis of chained arguments such as the Sorites. (Keep in mind that we have derivations on two levels now: the sequents themselves are derivations of a certain kind, but we also derive sequents from other sequents by the rules of the calculus.)

Consider again the first step of the Sorites chain. What we have there is in fact a composition of two derivations, $s = s_0, s_0 = s_1 \vdash s = s_1$ (by the transitivity of identity) and $M \vdash s_0 = s_1$ (by the elimination of the universal quantifier), composed by the cut rule as follows:

$$\frac{s = s_0, s_0 = s_1 \vdash s = s_1 \qquad M \vdash s_0 = s_1}{s = s_0, M \vdash s = s_1}\ \mathit{Cut}$$

A similar analysis applies to the second step, where cut is applied to compose $s = s_1, s_1 = s_2 \vdash s = s_2$ and $M \vdash s_1 = s_2$ into $s = s_1, M \vdash s = s_2$. The two steps are further composed by cut into

$$\dfrac{s = s_0, M \vdash s = s_1 \qquad s = s_1, M \vdash s = s_2}{s = s_0, M^2 \vdash s = s_2}\,\mathit{Cut}$$

where $M^2 = M, M$. In general, all steps in the inference are of the form $s = s_n, M \vdash s = s_{n+1}$ and can be brought together by cut to obtain a complete analysis of the Sorites chain:

$$\dfrac{\begin{array}{c}\dfrac{\dfrac{s = s_0, M \vdash s = s_1 \quad s = s_1, M \vdash s = s_2}{s = s_0, M^2 \vdash s = s_2}\,\mathit{Cut} \qquad s = s_2, M \vdash s = s_3}{s = s_0, M^3 \vdash s = s_3}\,\mathit{Cut} \\ \vdots \\ s = s_0, M^{999} \vdash s = s_{999}\end{array} \qquad s = s_{999}, M \vdash s = s_{1,000}}{s = s_0, M^{1,000} \vdash s = s_{1,000}}\,\mathit{Cut}$$

The natural *denotational interpretation* for the endsequent is

$$(\|s = s_0\| *_{\text{Ł}} \|M\|^{1,000}) \Rightarrow_{\text{Ł}} \|s = s_{1,000}\| \tag{3.2}$$

where $\|M\|^{1,000} = \|M\| *_{\text{Ł}} \ldots *_{\text{Ł}} \|M\|$. The sequent is valid (has the value 1) iff 3.1 holds (cf. Negri & von Plato, 2001, pp. 47, 58; Běhounek et al., 2011, § 4.3). The semantic and syntactic approaches to the paradox are obviously equivalent in this sense. But there is more to the story. In sequent calculi for classical logic, we have the rule of contraction,

$$\dfrac{A, A, \Gamma \vdash \Delta}{A, \Gamma \vdash \Delta}\,\mathit{Ctr}$$

If this rule were available in Łukasiewicz logic, we could apply it 999 times to derive $s = s_0, M \vdash s = s_{1,000}$—but it is not. Copeland (1997, pp. 522–525) concludes that it is the failure of contraction that is to be blamed for the Sorites paradox, rather than the failure of modus ponens (or in this case, the transitivity of identity). In our semantic framework, the explanation for this is of course the non-idempotence of the Łukasiewicz t-norm. If Copeland is right, then the paradox arises because we mistakenly apply contraction in a context where it is not applicable.

That is a neat story, but it is all too obvious that Copeland's analysis only applies if all identities $s_n = s_{n+1}$ are derived from the shared premiss M. Neither from a philosophical nor logical point of view is it necessary to base the argument on M (cf. Priest, 1991, pp. 293–294). We could just as well make a separate assumption for each identity (which is in fact what Priest does). The resulting paradox is essentially the same, and Priest's analysis applies equally well, but Copeland's analysis no longer does.[5] In Priest's analysis, we end up with

$$\|s = s_0\| *_{Ł} \|s_0 = s_1\| *_{Ł} \ldots *_{Ł} \|s_{999} = s_{1,000}\| \leq \|s = s_{1,000}\| \tag{3.3}$$

which yields the same value of $\delta_{1,000}$ as before. Our earlier remarks on the (non-)idempotence of the conjunction are no longer relevant, but that is not essential. What is essential is that the corresponding endsequent in Copeland's analysis becomes

$$s = s_0, s_0 = s_1, \ldots, s_{999} = s_{1,000} \vdash s = s_{1,000} \tag{3.4}$$

where each formula appears only once. The failure of contraction can no longer be offered as an explanation for the paradox, as there is nothing in 3.4 to contract. Copeland's solution is a solution to only one formulation of the paradox, which could be taken to show that it is not a real solution at all. What are we to blame for the paradox, then, if neither transitivity (modus ponens) nor contraction can be held responsible? What we have is a valid derivation of a valid endsequent from valid topsequents. If we can point to nothing to explain why the paradox arises, we have not really solved the paradox.

Perhaps the paradox is due to the denotational interpretation of sequents. On the classical interpretation, the endsequent 3.4 is valid iff

$$\|s = s_0\| \wedge \|s_0 = s_1\| \wedge \cdots \wedge \|s_{999} = s_{1,000}\| \leq \|s = s_{1,000}\|$$

which holds for crisp identity but not for fuzzy identity. Therefore, 3.4 is invalid on the classical interpretation and valid on the fuzzy one. Could the paradox be a confusion between the two interpretations? The paradox would then be an illusion that arises because we fail to recognize the proper weight of the assumptions that we have made. This would bring the syntactic approach very close to Priest's explanation, according to which the argument is globally invalid but locally valid. While Priest still considers it invalid and does not count local validity as a proper kind of validity, the syntactic approach gives no reason for such preference. On the contrary, it would seem as if local validity could be considered a kind of validity like any other.

---

[5] Note that our earlier remarks (p. 22) on the reducibility of the standard Sorites to the identity Sorites apply equally well without the shared major premiss.

3.4 Criticism from within the fuzzy framework

Priest's fuzzy identity can be criticized on the grounds that it is fuzzy. If identity is a crisp relation, then Priest's theory is not really a theory of *identity*. We will we take up this objection in Chapter 5, but first we have to ask whether Priest's theory is an adequate theory in the first place. Even if crispness were not a necessary condition for identity, we must still ask whether Priest's proposal meets the other conditions for identity, whatever those conditions are.

It has already been acknowledged that while Priest's identity does validate Ref, Sym and Tr& in Łukasiewicz logic, it validates SI& only for smooth predicates. Smooth predicates are defined by the condition $|f_P(x) - f_P(y)| \leq d(x, y)$, which is heavily dependent on the distance metric d, just like the definition of fuzzy identity in general. Fuzzy identity is thereby reduced to another concept, distance. It may be asked if this is any progress at all. Of course, distance is a better-understood notion in the sense that it has been studied within the mathematical theory of metric spaces, but that does not yet mean that it helps to make sense of fuzzy identity.

Besides giving the general mathematical definition (Definition 3.1.2), Priest does not characterize distance metrics in any other way than by giving examples of distance metrics for real numbers (e.g. $|x - y|$). He does not specify any distance metric for vague objects in the real world, and yet he explicitly suggests that his theory would be a theory of such objects (Priest, 2008, pp. 576–577). He motivates the theory with a Sorites paradox about a car that has its parts changed (Priest, 2008, pp. 572–573), but he does not give any distance metric for cars. In solving our version of the same paradox, we relied on the explicit definition $d(s_n, s_m) = \frac{|n-m|}{1{,}000}$, but this can hardly be considered anything else than an *ad hoc* function that was postulated for no other reason than to illustrate a general approach to solving the paradox.

Because of the mystery about the distance metric d, we cannot consider the requirement of smoothness very informative. A predicate validates SI& only if it smooth, but in order to know whether a predicate is smooth, we must first know the distance metric in question. On the other hand, even if we knew the distance metric in question, so what? If we are to classify Theseus's ship as a vague object with fuzzy identity, it is not sufficient to define the identity on a distance metric and point out that LL holds for predicates that are smooth on that distance metric. We must also know whether the predicates, or rather, *properties* of the ship expressed by those predicates are smooth, or else we cannot know whether LL applies to the identity of the ship. (In the context of real objects, it is more natural to consider LL instead of SI&, although they are logically equivalent.) Consider the identity $s_0 = s_{500}$, for instance: according to our distance metric, its truth value is 0.5. It follows that LL holds (without restriction) only if $|f_P(s_0) - f_P(s_{500})| \leq 0.5$ for all properties f_P. Can this requirement really be met? It is sufficient to name just one crisp property that is possessed by s_0 but not by s_{500}, or vice versa.

To define an identity for vague objects, would it not be better to start from the properties of those objects and define a distance metric so as to make them smooth? Priest does just the opposite: he starts from the distance metric, admits that the identity so defined does not generally validate LL, and then tries to remedy the situation by the observation that LL nevertheless holds for smooth predicates, which he regards as the kind of predicates that naturally pertains to vague objects. But here he is mistaken, I think. While it might be natural to expect vague objects to have fuzzy properties, fuzziness does not imply smoothness. Even the requirement that the properties be continuous or tolerant, in the ordinary sense, is too weak (cf. Priest, 1998, § 6; Priest, 2008, § 25.6.11). It is actually misleading of Priest to speak of smoothness as if it were a stronger version of the tolerance characteristic of fuzzy properties, for distance metrics and smooth properties as such have nothing to do with fuzziness or tolerance (for instance, the complement of classical identity is a distance metric that makes all properties trivially smooth). From a mathematical point of view, the smoothness condition is exactly the condition that a predicate must meet in order to validate SI& on a given distance metric (cf. p. 30). Priest's observation that LL holds for smooth predicates amounts to just saying that LL holds for predicates for which it holds. He introduces smooth predicates as if they were to yield a fuzzy version of LL, but he fails to provide them with an independent motivation. It all depends on the distance metric, and the general definition of a distance metric alone does not guarantee any compliance with Leibniz's Law; nor does the more specific definition employed in our solution to the identity Sorites.

The sacrifice of SI& has profound consequences for our attempt to show that the identity Sorites could be solved analogically with the standard Sorites. In his first paper on the two Sorites, Priest (1991) writes the identity Sorites in the general form (SSD for monadic predicates) with the rule of SI, argues that it be interderivable with the standard Sorites, and suggests that the same kind of solution would apply to both kinds of Sorites. However, when he later (Priest, 1998, 2008) proposes a solution to the identity Sorites, he does not give a solution in the general case but only in the special case with transitivity (SD), which makes a big difference for his theory. The predicate that is transferred there by transitivity is identity itself, which is smooth by definition. In the general case, by contrast, there is an arbitrary predicate that is transferred by SI, which is locally valid only if the predicate is smooth. There is, in this respect, a certain disparity between Priest's original vision and his eventual theory. We have seen (in Section 2.3) that both the standard Sorites (SS) and the special identity Sorites (SD) are reducible to the general form (SSD), and yet they are treated by Priest as two analogous but separate paradoxes with analogous but separate solutions. Given that the two paradoxes can be combined into one, why cannot their solutions be so combined too? Because SI is not locally valid in Priest's logic.

The original question was whether fuzzy identity meets the conditions for identity other than crispness, "whatever those conditions are." It is time to ask, what exactly are those conditions? According to Noonan and Curtis (2018, § 2), the "classical view" of identity characterizes it as "the equivalence relation which everything has to itself and to nothing else and which satisfies Leibniz's Law." How exactly these requirements should be translated into fuzzy logic is a matter that will not be settled here; such a translation would no doubt force us to step back and confront some fundamental questions about what 'identity' actually means in philosophy, the discussion of which will be postponed until Section 5.4. For now, however, I think it safe to say that it would be a great plus for a candidate for fuzzy identity if it at least complied with the fuzzy versions of the classical principles about identity, such as Ref, Sym, Tr&, and LL; but, as we have seen, Priest's fuzzy identity does not satisfy LL in Łukasiewicz logic—nor does it in any fuzzy logic, for that matter—and that is the end of the story.

On a higher level, we might say that Priest's theory rests on the fundamental assumption that we can redefine identity by introducing a new notion from outside the traditional theory of identity, even if that condition fails to preserve what was traditional about identity. We might go as far as to suggest that it is this fundamental assumption which makes his fuzzy identity not really identity. There is an alternative way to define fuzzy identity that makes no such assumption, in a way that arises naturally from inside the traditional theory of identity. The solution lies within the traditional principles themselves, and Leibniz's Law in particular.

Chapter 4

Fuzzy identity reconsidered

In this chapter, we will refine Priest's fuzzy identity into a form that better appreciates the traditional principles of identity. We will proceed as with the presentation of Priest's original theory in Chapter 3, providing first some mathematical background (Sections 4.1 and 4.2) against which the theory of fuzzy identity will then be reformulated (Sections 4.3 and 4.4). The theory will be critically discussed from a philosophical perspective in Chapter 5.

4.1 Representation theorem

We will continue the overview of the theory of similarity relations that was started in Section 3.1. We will introduce the notion of an extensional set, which is a generalization of Priest's notion of a smooth property, and study its relationship with the notion of a similarity relation.

Let us denote by $[0,1]^D$ the set of all functions $D \longrightarrow [0,1]$, i.e., the set of all (fuzzy) subsets of the domain D.

Definition 4.1.1 (Extensionality)**.** Let $F \in [0,1]^D$, and S a $*$-similarity relation on D. The set F is *extensional* with respect to S if

$$F(x) * S(x,y) \leq F(y)$$

for all $x, y \in D$.

Obviously, extensional sets are precisely those properties that validate SI& (or LL) in a fuzzy logic with fuzzy identity, where $I(=) = S$. To deal with sets that are not extensional (such as crisp equivalence classes), we can consider their smallest extensional supersets, called their *extensional hulls* (see Klawonn & Castro, 1995, pp. 204–205). That F is extensional with respect to S can be equivalently defined by the condition that F is a *generator* of S,

$$S(x,y) \leq F(x) \Leftrightarrow_* F(y)$$

for all $x, y \in D$ (Jacas & Recasens, 1992, Th. 4.2; Klawonn & Castro, 1995, Th. 2.10).

If S is a similarity relation, we will denote the set of all its generators by $\mathcal{F}_S$. It can be proved that each similarity relation S can in fact be represented in a form where it is literally generated by some $\mathcal{F} \subseteq \mathcal{F}_S$. Conversely, every relation represented in such a form is a similarity relation. This result has been proved by Valverde (1985, Th. 4.2):

Theorem 4.1.2 (Representation theorem). *Let S be a binary relation on the domain D, and $*$ a continuous t-norm. Then S is a $*$-similarity relation iff there is a family of sets $\mathcal{F} \subseteq [0,1]^D$ such that*

$$S(x,y) = \inf_{F \in \mathcal{F}} \{F(x) \Leftrightarrow_* F(y)\} \tag{RT}$$

for all $x, y \in D$.

If S is determined by $\mathcal{F}$ through RT, we write $S = S_{\mathcal{F}}$ and say that $S_{\mathcal{F}}$ is RT-generated by $\mathcal{F}$, or that $\mathcal{F}$ is a set of RT-generators of S. It can be proved that $S_{\mathcal{F}}$ is the coarsest (greatest) similarity relation generated by $\mathcal{F}$, i.e., such that all sets in $\mathcal{F}$ are extensional with respect to S (Klawonn & Castro, 1995, Th. 3.1).

Contrast RT with the fuzzy truth conditions of LL (the indiscernibility of identicals) and II (the identity of indiscernibles),

$$\begin{aligned} \|x = y\| &\Rightarrow_* \inf_{A \in \mathcal{A}} \|A(x) \leftrightarrow A(y)\| \\ \inf_{A \in \mathcal{A}} \|A(x) \leftrightarrow A(y)\| &\Rightarrow_* \|x = y\| \end{aligned}$$

where $\mathcal{A}$ is the set of "all formulae A." If identity is interpreted as similarity, the requirement that LL and II be fully true translates into

$$S(x,y) \leq \inf_{F \in \mathcal{F}} \{F(x) \Leftrightarrow_* F(y)\} \tag{LL*}$$

$$\inf_{F \in \mathcal{F}} \{F(x) \Leftrightarrow_* F(y)\} \leq S(x,y) \tag{II*}$$

where $\mathcal{F}$ is the set of properties on (or subsets of) D defined by $\mathcal{A}$. It is now obvious that RT is the conjunction of LL* and II*. The former sets an upper bound and the latter a lower bound on the value of $S(x,y)$. Hence, every similarity relation S obeys both LL and II with respect to a set of its RT-generators.

Note that a set of RT-generators of S is not unique: the same S may be RT-generated by distinct $\mathcal{F}$s, which can be shown by a simple example.[1] However, it can be proved that S is the

---

[1] Let e.g. $D = \{a, b\}, \mathcal{F} = \{f\}, \mathcal{F}' = \{f'\}$, where $f(a) = f'(b) = 1$ and $f(b) = f'(a) = 0$.

only similarity relation with the set of generators $\mathcal{F}_S$, i.e., S is uniquely determined (through RT) by $\mathcal{F}_S$ (see Klawonn & Castro, 1995, p. 210).

This disparity between a set of RT-generators and the set of all generators of a similarity relation S raises a question about the smallest subset $\mathcal{F} \subseteq \mathcal{F}_S$ that is sufficient to RT-generate S. This question about 'minimal generators' has been studied by Jacas (1990). As for the set $\mathcal{F}_S$ of 'maximal generators,' it has been shown that not only is $\mathcal{F}_S$ closed under certain properties, but also every set that is closed under them is the set of generators of a (unique) similarity relation.

Theorem 4.1.3 (Klawonn & Castro, 1995, Th. 3.2, 3.3)**.** *The set $\Phi \subseteq [0,1]^D$ is the set of generators of a similarity relation on D iff for all $\phi \in \Phi, \Psi \subseteq \Phi, \alpha \in [0,1]$,*

1. $(\bigcup \Psi) \in \Phi$
2. $(\bigcap \Psi) \in \Phi$
3. $(\alpha * \phi) \in \Phi$
4. $(\phi \Rightarrow_* \alpha) \in \Phi$
5. $(\alpha \Rightarrow_* \phi) \in \Phi$

If the underlying algebra is $[0,1]_{\text{Ł}}$, it is sufficient that Φ meet conditions 1 and 4 (Klawonn & Castro, 1995, Th. 3.4).

4.2 'Local extensionality'

The results of the previous section give new content to Priest's requirement of smooth predicates. If d is a distance metric and S is the Łukasiewicz similarity relation defined by $S = 1 - d$, the smooth properties on d are exactly the extensional sets with respect to S, or the generators $F \in \mathcal{F}_S$ of S. What is more, the set $\mathcal{F}_S$ is closed with respect to the conditions 1–5 of Theorem 4.1.3, which helps us to find formulae A such that the inference from $A(x)$ and $x = y$ to $A(y)$ by the rule of SI is locally valid.

There is more to be said on this matter, for the local validity of SI does not require that the property defined by A be included in the set $\mathcal{F}_{I(=)}$; it merely requires that the drop in truth value from the premisses to the conclusion be bounded, so that the conclusion can be brought arbitrarily close to full truth by making the premisses true enough. It turns out that the drop can be bounded for all formulae whose non-logical atoms are in the set of (RT-)generators.

Definition 4.2.1. Let $A(x)$ be a formula, $\mathcal{M} = (D, I)$ a model, v an assignment from variables to D, $f^{\mathcal{M},v}_{A(x)} \in [0,1]^D$ the set defined by $A(x)$,[2] and $\Phi \subseteq [0,1]^D$. We say that $A(x)$ is a Φ-*derived*

---

[2] Recall the interpretation of formulae as sets defined on p. 11.

formula under $\mathcal{M}$ and v if some of the following conditions hold:

- $f^{\mathcal{M},v}_{A(x)} \in \Phi$;
- $A(x)$ is an identity formula;
- $A(x)$ is a compound of Φ-derived formulae.

In other words, $A(x)$ is *not* a Φ-derived formula only if it contains an atomic formula which is neither an identity formula nor included in Φ.

Definition 4.2.2. Let $\mathcal{M} = (D, I)$ be a model, v an assignment from variables to D, and $\Phi \subseteq [0,1]^D$. The Φ-*weight* of a Φ-derived formula $A(x)$ under $\mathcal{M}$ and v, $w(A)$, is defined as $w(A) = 1$ whenever $f^{\mathcal{M},v}_{A(x)} \in \Phi$ or A is an identity formula. Otherwise,

$$\begin{array}{ll} w(A) = 0 & \text{for } A = \bot \\ w(A) = w(B) & \text{for } A = \forall x B, \exists x B \\ w(A) = w(B) \vee w(C) & \text{for } A = (B \wedge C), (B \vee C) \\ w(A) = w(B) + w(C) & \text{for } A = (B \mathbin{\&} C), (B \to C), \text{ where } \& \neq \wedge \end{array}$$

Weights for the remaining connectives are derived as follows:

$$\begin{aligned} w(\top) &= w(\bot \to \bot) = w(\bot) + w(\bot) = 0 \\ w(\neg B) &= w(B \to \bot) = w(B) + w(\bot) = w(B) \\ w(B \leftrightarrow C) &= w((B \to C) \wedge (C \to B)) = w(B) + w(C) \end{aligned}$$

The motivation behind this definition will reveal itself in the following theorem. We will simplify the notation by substituting α, β, γ for $f^{\mathcal{M},v}_{A(x)}, f^{\mathcal{M},v}_{B(x)}, f^{\mathcal{M},v}_{C(x)}$.

Theorem 4.2.3. *Let $\mathcal{M} = (D, I)$ be a model, v a variable assignment to D, S a similarity relation on D, Φ a set of RT-generators of S, $A(x)$ a Φ-derived formula under $\mathcal{M}$ and v, and $w(A)$ the Φ-weight of $A(x)$. Then for all $x, y \in D$,*

$$\alpha(x) * S(x,y)^{w(A)} \leq \alpha(y)$$

where $\alpha(x) = f^{\mathcal{M},v}_{A(x)}(x)$.

Proof. By induction on the definition of $w(A)$.

If $\alpha \in \Phi$, then $w(A) = 1$ and the claim holds by the definition of Φ.

If A is an identity formula, then $w(A) = 1$ and the claim holds by the $*$-transitivity of S.

If $A = \bot$, then $\alpha(x) = \alpha(y) = 0$ for all $x, y \in D$ and the claim holds trivially.

If $A = \forall z B$, we have two subcases. If x is not free in A, then α is a constant set as in the previous case. If x is free in A, then it must be free in B, and z may be free in B, too. Since the set defined by $B(x)$ depends on the value assigned to z, we must adjust our notation to provide access to that value. Let $\beta_a(x) = f_{B(x)}^{\mathcal{M},v(a/z)}(x)$ for each $a \in D$. It follows that $\alpha(x) = \inf_{a \in D} \beta_a(x)$, and by inductive hypothesis

$$\beta_a(x) * S(x,y)^{w(B)} \leq \beta_a(y)$$

from which we infer

$$\inf_{a \in D} \beta_a(x) * S(x,y)^{w(B)} \leq \inf_{a \in D} \beta_a(y)$$

If $A = \exists z B$, we reason as in the previous case.

If $A = B \wedge C$, then $\alpha(x) = \beta(x) \wedge \gamma(x)$ and by inductive hypothesis

$$\beta(x) * S(x,y)^{w(B)} \leq \beta(y)$$
$$\gamma(x) * S(x,y)^{w(C)} \leq \gamma(y)$$

whence it follows by the monotonicity of $\wedge$

$$\left(\beta(x) * S(x,y)^{w(B)}\right) \wedge \left(\gamma(x) * S(x,y)^{w(C)}\right) \leq \beta(y) \wedge \gamma(y)$$

By the monotonicity of $*$, we know that both $\beta(x) * S(x,y)^{w(B)}$ and $\gamma(x) * S(x,y)^{w(C)}$ are greater than or equal to $(\beta(x) \wedge \gamma(x)) * \left(S(x,y)^{w(B)} \wedge S(x,y)^{w(C)}\right)$, and therefore

$$(\beta(x) \wedge \gamma(x)) * \left(S(x,y)^{w(B)} \wedge S(x,y)^{w(C)}\right) \leq \beta(y) \wedge \gamma(y) \qquad \Longrightarrow$$
$$(\beta(x) \wedge \gamma(x)) * S(x,y)^{w(B) \vee w(C)} \leq \beta(y) \wedge \gamma(y)$$

If $A = B \vee C$, then by decreasing the left-hand sides and increasing the right-hand sides of the inductive hypothesis,

$$\beta(x) * S(x,y)^{w(B) \vee w(C)} \leq \beta(y) \vee \gamma(y)$$
$$\gamma(x) * S(x,y)^{w(B) \vee w(C)} \leq \beta(y) \vee \gamma(y)$$

whence

$$(\beta(x) \vee \gamma(x)) * S(x,y)^{w(B) \vee w(C)} \leq \beta(y) \vee \gamma(y)$$

If $A = B \mathbin{\&} C$, then by inductive hypothesis and the monotonicity of $*$

$$(\beta(x) * \gamma(x)) * S(x,y)^{w(B)+w(C)} \leq \beta(y) * \gamma(y)$$

If $A = B \rightarrow C$, then by the definition of $\Rightarrow_*$,

$$\beta(x) * (\beta(x) \Rightarrow_* \gamma(x)) \leq \gamma(x)$$

By the monotonicity of $*$,

$$\beta(x) * (\beta(x) \Rightarrow_* \gamma(x)) * S(x,y)^{w(C)} \leq \gamma(x) * S(x,y)^{w(C)}$$

By applying the inductive hypothesis on both sides, we get

$$\beta(y) * S(y,x)^{w(B)} * (\beta(x) \Rightarrow_* \gamma(x)) * S(x,y)^{w(C)} \leq \gamma(y)$$

whence by RC, the symmetry of S, and the commutativity of $*$,

$$(\beta(x) \Rightarrow_* \gamma(x)) * S(x,y)^{w(B)+w(C)} \leq \beta(y) \Rightarrow_* \gamma(y) \qquad \square$$

We might call this result 'local extensionality,' for it extends the notion of extensionality to very much the same direction as Priest's local validity extends the notion of validity. The connection between local extensionality and local validity will be made clear in Theorem 4.4.2.

Note that in Gödel logic,

$$\alpha(x) * S(x,y) \leq \alpha(x) * S(x,y)^{w(A)} \leq \alpha(y)$$

i.e., local extensionality reduces back to ordinary extensionality. This is because the Gödel t-norm is *idempotent*: $x \wedge x = x$ for all $x \in [0,1]$. As desirable as this feature may seem, we will soon discover other features of Gödel logic that make it altogether incompatible with what we are trying to accomplish here.

4.3 Varieties of similarity

We have seen that every $*$-similarity relation S can be represented in the form

$$S(x, y) = \inf_{\phi \in \Phi} \{\phi(x) \Leftrightarrow_* \phi(y)\}$$

As Klawonn and Castro (1995, p. 208) point out, $\phi(x) \Leftrightarrow_* \phi(y)$ can be interpreted as the degree to which x and y cannot be distinguished by the set $\phi \in \Phi$, and $\inf_{\phi\in\Phi}\{\phi(x) \Leftrightarrow_* \phi(y)\}$ as the degree to which x and y cannot be distinguished by the family Φ of sets. For this reason, it is natural to think of a similarity relation as *indistinguishability* with respect to a given set of properties—or *indiscernibility*, if one wishes to underline the connection with LL* and II*.

We introduce the notation $Z_{*,\Phi}$ to denote a $*$-similarity relation RT-generated by Φ. (We may omit one of $*, \Phi$ when the other is known or irrelevant.) The new notation emphasizes the dependability of a similarity relation on the associated t-norm, on one hand, and on the properties with respect to which the similarity is evaluated, on the other hand. It is worth digging a bit deeper into this dependability, and the multitude of different similarity relations, so as to be able to judge the acceptability of similarity as a kind of *identity*.

Theorem 4.3.1. *Let Z_* and $Z_{*'}$ be similarity relations for t-norms $*$ and $*'$ for a common family Φ on a domain D. If $* \leq *'$, then for all $a, b \in D$,*

$$Z_*(a, b) \geq Z_{*'}(a, b)$$

Proof. Let $x, y \in [0, 1]$. Suppose $x > y$. Obviously, $y \Rightarrow_* x = 1 = y \Rightarrow_{*'} x$. By hypothesis, $x * z \leq x *' z$ for all $z \in [0, 1]$ and therefore

$$x \Rightarrow_* y = \sup\{z \mid x * z \leq y\} \geq \sup\{z \mid x *' z \leq y\} = x \Rightarrow_{*'} y$$

It follows that

$$\min\{x \Rightarrow_* y, y \Rightarrow_* x\} \geq \min\{x \Rightarrow_{*'} y, y \Rightarrow_{*'} x\}$$

and likewise if $x \leq y$. We conclude that $x \Leftrightarrow_* y \geq x \Leftrightarrow_{*'} y$ for all x, y; specifically,

$$\inf_{\phi\in\Phi}\{\phi(a) \Leftrightarrow_* \phi(b)\} \geq \inf_{\phi\in\Phi}\{\phi(a) \Leftrightarrow_{*'} \phi(b)\} \qquad \square$$

This is in harmony with our earlier remarks on the mutual inclusion of similarity relations (Figure 3.1). The larger the underlying t-norm, the closer the relation is to classical identity.

Gödel similarity is closer to classical identity than product similarity, which in turn is closer than Łukasiewicz similarity. Yet it would be an error to conclude that Gödel similarity is the best candidate for fuzzy identity; on the contrary, it is utterly unsuitable for the job. To begin with, material biconditional in Gödel logic is equivalent to conjunction. This is suspicious enough in itself, but it has even more absurd consequences when biconditional is employed in the generation of similarity. Rather than a genuine comparison of x and y in some respect, the similarity $Z_{G,\Phi}(x, y)$ is in effect just an infimum of $\phi(z)$, where $\phi \in \Phi$ and $z \in \{x, y\}$. It follows e.g. that the identity $x = y$ is fully false whenever both x and y fully lack some property, which is absurd; if anything, the fact that x and y possess some property to exactly the same degree should be evidence *for* their identity, and certainly not against it.

The biconditional of product logic (recall Table 1.1) behaves better, but there are still some peculiarities when it comes to similarity. Suppose that x and y are indiscernible with respect to all properties in Φ except ϕ, for which the difference $|\phi(x) - \phi(y)|$ is negligible; say, 0.001. Then the value of $x = y$ approaches 0 when $\phi(x)$ and $\phi(y)$ approach 0, although the difference $|\phi(x) - \phi(y)|$ is kept fixed. If $\phi(x) = 0.999$ and $\phi(y) = 0.998$, the value of $x = y$ is $0.9989\ldots$, but if $\phi(x) = 0.0001$ and $\phi(y) = 0.0011$, the value of $x = y$ is $0.0909\ldots$. On the other hand, the value of $x = y$ is 0.5 when e.g. $\phi(x) = 0.9, \phi(y) = 0.45$, but it has the same value when $\phi(x) = 0.001, \phi(y) = 0.002$, although the difference $|\phi(x) - \phi(y)|$ is remarkably higher in the former case. If that is not counterintuitive enough, just replace ϕ in Φ with its negation; then $\neg_\Pi \phi(x) = \neg_\Pi \phi(y) = 0$ and the value of $x = y$ becomes 1 in all cases.

Regardless of what properties are included in RT-generators, it is controversial whether product similarity is applicable to the philosophical problem at hand. If it is, then we must think of fuzzy identity in a completely different way than what appears to be the intuitive interpretation.[3] If it is not, then Łukasiewicz similarity is the only prominent candidate for the position of fuzzy identity, although it is farther from classical identity (in the sense of Theorem 4.3.1) than the other two. In any case, as Priest's definition of fuzzy identity is indeed based on the Łukasiewicz t-norm, it is not his choice of t-norm that should be held responsible for the problems of his theory; the solution must be sought elsewhere.

Next, let us see what happens when the set of RT-generators is increased while the t-norm is kept fixed.

Theorem 4.3.2. *Let Z_Φ and $Z_{\Phi'}$ be similarity relations for families Φ and Φ' for a common t-norm $*$ on*

---

[3] For an early treatment of vague concepts in terms of product logic, see Goguen, 1968–1969.

a domain D. If $\Phi \subseteq \Phi'$, then for all $a, b \in D$,

$$Z_{\Phi}(a,b) \geq Z_{\Phi'}(a,b)$$

Proof. If for all $\varphi' \in \Phi \cap \Phi'$,

$$\inf_{\phi \in \Phi} \{\phi(a) \Leftrightarrow_* \phi(b)\} \leq \varphi'(a) \Leftrightarrow_* \varphi'(b)$$

then

$$\inf_{\phi \in \Phi} \{\phi(a) \Leftrightarrow_* \phi(b)\} = \inf_{\phi \in \Phi'} \{\phi(a) \Leftrightarrow_* \phi(b)\}$$

and the claim follows trivially; else there is some $\varphi' \in \Phi \cap \Phi'$ such that

$$\inf_{\phi \in \Phi} \{\phi(a) \Leftrightarrow_* \phi(b)\} > \varphi'(a) \Leftrightarrow_* \varphi'(b)$$

in which case

$$\inf_{\phi \in \Phi} \{\phi(a) \Leftrightarrow_* \phi(b)\} > \inf_{\phi \in \Phi'} \{\phi(a) \Leftrightarrow_* \phi(b)\}$$

and the claim follows. □

We thus obtain a new hierarchy of similarity relations, complementary to the one presented earlier. The inclusion relation $\subseteq$ on the sets $\Phi \in [0,1]^D$ of RT-generators induces a partial ordering on the respective similarity relations Z_{Φ}. For the smallest subset of $[0,1]^D$—the empty set—we obtain as a limiting case of RT a similarity relation $Z_{\emptyset}$ such that $Z_{\emptyset}(x,y) = 1$ for all $x, y \in D$, i.e., all objects are completely indiscernible from each other. As for the largest subset of $[0,1]^D$—i.e., $[0,1]^D$ itself—it can easily be confirmed that the corresponding similarity reduces to classical identity, whereby all objects are completely discernible from each other. In between the two extremes, we have a continuum of similarity relations of increasing strength.

It follows that any similarity relation can be gradually strengthened by adding more functions to the set of RT-generators until it finally reduces to classical identity, the strongest similarity of all. Specifically, if we add the crisp characteristic functions of (the equivalence classes of) all individual elements $a \in D$,

$$f_a(x) = \begin{cases} 1, & \text{for } x = a \\ 0, & \text{for } x \neq a \end{cases}$$

then the resulting similarity relation is the classical identity relation on D irrespective of its other generators, if any (cf. the discussion on circularity on p. 15).

4.4 Beginnings of a new logic

What are we to say about Priest's theory of fuzzy identity in light of the new developments?

Recall that in Section 3.4, we criticized the theory for not being specific enough about the nature of the underlying distance metric and for not generally complying with LL. Priest's attempt to alleviate this shortage by the introduction of smooth properties was accused of begging the question, for he effectively just defined what it is for a property to comply with LL under some unspecified distance metric. We have now seen that there is no need to maintain any mystery about the distance metric, for any similarity relation can be generated by (some of) its extensional properties, which correspond exactly to Priest's smooth properties in Łukasiewicz logic. While we still can extract a distance metric from the representation,[4] it no longer plays a significant role as it is completely reduced to the generators. It is therefore perfectly possible, and plausibly preferable from a philosophical point of view, to base the theory of fuzzy identity not on the notion of a distance metric but on those properties that generate the underlying similarity relation.

What exactly are the generating properties is a matter of metaphysics. To some extent, the choice is arbitrary, for the same similarity relation can be generated by different selections of properties. If the generating properties are all crisp, then the relation is also crisp; if they are all fuzzy, so is the relation (at least for a finite number of properties); and if the properties include both crisp and fuzzy ones, the resulting relation may take both crisp and fuzzy values depending on the circumstances. In any case, the choice of the generators is left to the metaphysician; whatever properties are chosen, fuzzy identity can be logically constructed out them.

In order to demonstrate the logical aspects of fuzzy identity without committing to any particular set Φ of RT-generators, we will suppose, for the sake of simplicity, that the properties in Φ are defined by the non-logical atoms in a given language $\mathcal{L}$, i.e., atomic formulae of the form $R(t_1, \ldots, t_n)$. It then follows that all $\mathcal{L}$-formulae are Φ-derived. This appears to be a natural, logically convenient, and yet ontologically neutral choice: it is then up to the metaphysician to come up with the specific non-logical predicates, or the corresponding properties, taken to constitute fuzzy identity.

---

[4] The corresponding distance metric of a Łukasiewicz similarity relation $Z_{Ł,\Phi}$ reduces by RT to $d_Ł(x, y) = \sup_{\phi \in \Phi}\{|\phi(x) - \phi(y)|\}$.

Let $\mathcal{M} = (D, I)$ be a Ł-model of a language $\mathcal{L}$, v an associated assignment, and

$$\Phi = \left\{ f^{\mathcal{M},v}_{A(x)} \middle| A(x) \text{ is a non-logical atom in } \mathcal{L} \right\}$$

We can then redefine identity as

$$\|t_1 = t_2\|^{\mathcal{M}}_v = Z_{Ł,\Phi}({t_1}^{\mathcal{M}}_v, {t_2}^{\mathcal{M}}_v)$$

We thus obtain a new variant of Łukasiewicz logic with fuzzy identity. To distinguish it from Priest's version and to emphasize its (partial) internalization of LL and II, we shall call it *Łukasiewicz logic with* strong *fuzzy identity* and denote it by $Ł_Z$. The only thing that has changed from Priest's version is the specification of the RT-generators. Unlike Priest, who did not specify any RT-generators, we have fixed the set of RT-generators as the non-logical atoms of $\mathcal{L}$.

The above definition is the 'intended' interpretation of fuzzy identity in $Ł_Z$. The definition is a model-theoretical one, but it is based on the properties defined by the non-logical atoms of the first-order language $\mathcal{L}$. Alternatively, and perhaps more naturally, we could define fuzzy identity axiomatically as follows.

In Section 1.3, we discussed the impossibility of defining crisp identity in first-order logic. The closest one can get to the intended notion is a crisp equivalence relation $\sim$ by the axioms Ref, SI&/LL, and Cr (p. 15). Now, if we restrict the formulae $A(x)$ in SI&/LL to atomic ones[5] and drop Cr (which was redundant in Łukasiewicz logic anyway and must now be excluded to make room for fuzzy identity), then in any Ł-model $\mathcal{M}$ of the axioms, the identity predicate $=$ defines a fuzzy relation $\approx$ such that

$$\begin{aligned} &f_{\approx}(x^{\mathcal{M}}_v, x^{\mathcal{M}}_v) = 1 \\ &f_{\approx}(x^{\mathcal{M}}_v, y^{\mathcal{M}}_v) \leq \inf_{A(x)} \{ \|A(x)\|^{\mathcal{M}}_v \Leftrightarrow_{Ł} \|A(y)\|^{\mathcal{M}}_v \} \end{aligned}$$

where $A(x)$ is atomic, and y is free for x. Symmetry and transitivity follow as before, making $\approx$ a similarity relation. It is not unique, however: the conditions are satisfied not only by the intended fuzzy identity but also the crisp identity and everything in between the two. This leaves us with a similar dilemma as in defining crisp identity.[6] It is not necessarily a bad thing;

---

[5] This is superficially different from the model-theoretic definition, where the defining formulae are restricted to non-logical atoms only. Identity formulae cannot be included or the definition would be circular, but the difference is inessential since equivalence relations are necessarily extensional with respect to themselves.

[6] The intended identity would be uniquely defined if we could express II as a (schematic) axiom, but we cannot. To uniquely define crisp identity, by contrast, we should be able to quantify over all properties and not just formulae of the language.

it just reflects the fact that the theory of fuzzy identity is a generalization of that of crisp identity. All its theorems are true of crisp identity, but crisp identity is not the only relation they are true of. In this respect, the situation has not changed from the theory of crisp identity. What has changed is that the theory now also allows the interpretation of = as a fuzzy relation. Is this not precisely what we wanted to achieve?

Now that we have restricted ourselves to Łukasiewicz logic, the only difference between the axiomatizations of crisp and fuzzy identity is the latter's restriction on the formulae $A(x)$ in SI&/LL, and even this difference disappears when only classical truth values $\{0,1\}$ are considered.[7] Yet the latter axioms are not optimal for proof-theoretic purposes precisely because of this restriction. It is time to return to the idea of 'local extensionality' suggested in Section 4.2 and see how it can be put to use in reasoning with fuzzy identity.

Lemma 4.4.1. *If $x \in [0,1]$ and $n \in \mathbb{N}$, then*

$$x^n = \max\{nx - n + 1, 0\}$$

Proof. By the observation that

$$x^n = \max\{\overbrace{x + x + x + \cdots + x}^{n} \underbrace{-1 - 1 - \cdots - 1}_{n-1}, 0\}$$

Since $x *_{Ł} x \leq x$, it is sufficient to perform the maximization with 0 at the final stage only. □

Theorem 4.4.2. *The rule of substitutivity of identicals (SI),*

$$A(x), x = y \vdash A(y)$$

is locally valid in $Ł_Z$ for all formulae A.

Proof. Let $\mathcal{M}$ be a model, v an assignment, $t = x_v^{\mathcal{M}}$ and $u = y_v^{\mathcal{M}}$. Let us abbreviate

$$\|A(x)\|_v^{\mathcal{M}} = f_{A(x)}^{\mathcal{M},v}(t) = B \qquad \|x = y\|_v^{\mathcal{M}} = Z_{Ł,\Phi}(t,u) = Q$$
$$\|A(y)\|_v^{\mathcal{M}} = f_{A(x)}^{\mathcal{M},v}(u) = C \qquad w(A) = w$$

Let ε be the level of acceptability required of the conclusion C. From the definition of Φ as the set of non-logical atoms, it follows that $A(x)$ is a Φ-derived formula. Hence it follows by

---

[7] See e.g. Negri & von Plato, 2011, pp. 107–108.

Theorem 4.2.3 that $B *_{Ł} Q^w \leq C$, and by Lemma 4.4.1,

$$B + wQ - w \leq B + Q^w - 1 \leq B *_{Ł} Q^w$$

If $w = 0$, the inference is trivially (globally) valid (cf. Definition 4.2.2 and Theorem 4.2.3). Suppose $w > 0$. Then

$$B + Q - 1 \leq B + wQ - w$$

and we can set $\varepsilon < B + Q - 1$, whence $\delta = \frac{\varepsilon+1}{2}$. □

We can similarly show that the rule of LL is locally valid for all formulae. We can also generalize the laws of SI& and LL into

$$\left(\mathop{\&}_{w(A)}(x = y) \mathbin{\&} A(x)\right) \to A(y) \tag{SI\&^{w}}$$

$$\mathop{\&}_{w(A)}(x = y) \to (A(x) \to A(y)) \tag{LL\&^{w}}$$

which are valid for all formulae $A(x)$, where y is free for x. We have thus discovered a fuzzy logic with fuzzy identity to which Priest's notion of local validity applies even better than to his original logic.

Example 4.4.3. The following sentences are theorems of $Ł_Z$ for all terms x, y, z, predicate symbols P and function symbols f:

$$(x = y \mathbin{\&} P(x)) \to P(y) \qquad (x = y \mathbin{\&} \neg x = z) \to \neg y = z$$

$$x = y \to (P(x) \leftrightarrow P(y)) \qquad \neg x = y \to (\neg x = z \oplus \neg y = z)$$

$$(P(x) \mathbin{\&} \neg P(y)) \to \neg x = y \qquad x = y \to f(x) = f(y)$$

where $B \oplus C \equiv \neg(\neg B \mathbin{\&} \neg C)$ is the strong disjunction (t-conorm) dual to & ($*_{Ł}$).

Example 4.4.4. We will consider a very simple set theory in the language $\mathcal{L} = \{a, b, q, r, \in\}$, where a, b are intended to denote individuals and q, r fuzzy sets thereof, and $\in$ is a binary predicate for the relation "is a member of." Let $\mathcal{M} = (D, I)$ be a model of $\mathcal{L}$, where $D = \{a^{\mathcal{M}}, b^{\mathcal{M}}, q^{\mathcal{M}}, r^{\mathcal{M}}\}$ and $I(\in)$ is a fuzzy relation with the characteristic function $f_{\in}$ such that

$$f_{\in}(a^{\mathcal{M}}, q^{\mathcal{M}}) = 0.3 \qquad f_{\in}(a^{\mathcal{M}}, r^{\mathcal{M}}) = 0.6$$

$$f_{\in}(b^{\mathcal{M}}, q^{\mathcal{M}}) = 0.9 \qquad f_{\in}(b^{\mathcal{M}}, r^{\mathcal{M}}) = 0.5$$

We may assume that the value of $f_\in(x, y)$ for any other (x, y) in D^2 is either 0 or undefined (the details do not matter here). It follows that

$$
\begin{aligned}
&|f_\in(a^\mathcal{M}, q^\mathcal{M}) - f_\in(b^\mathcal{M}, q^\mathcal{M})| = 0.6 \qquad && |f_\in(a^\mathcal{M}, r^\mathcal{M}) - f_\in(b^\mathcal{M}, r^\mathcal{M})| = 0.1 \\
&|f_\in(a^\mathcal{M}, q^\mathcal{M}) - f_\in(a^\mathcal{M}, r^\mathcal{M})| = 0.3 \qquad && |f_\in(b^\mathcal{M}, q^\mathcal{M}) - f_\in(b^\mathcal{M}, r^\mathcal{M})| = 0.4
\end{aligned}
$$

and therefore

$$
\|a = b\|^\mathcal{M} = \inf\{1 - 0.6, 1 - 0.1\} = 0.4 \qquad \|q = r\|^\mathcal{M} = \inf\{1 - 0.3, 1 - 0.4\} = 0.6
$$

The identity $a = b$ depends on the extent to which a and b are members of the same sets, and the identity $q = r$ on the extent to which q and r contain the same members. More formally, $a = b$ is the indiscernibility of a and b with respect to properties $Q(x) = x \in q$ and $R(x) = x \in r$, and $q = r$ is the indiscernibility of q and r with respect to properties $A(x) = a \in x$ and $B(x) = b \in x$. Interestingly, both $Q(x), R(x)$ and $A(x), B(x)$ define sets on the domain, i.e., both $\|a = b\|^\mathcal{M}$ and $\|q = r\|^\mathcal{M}$ depend on the extent to which the flanking terms are members of the same sets. Even though the identity of individuals and that of sets are different in the object language, they both reduce to the identity of individuals (members of the domain) in the metalanguage.

Finally, we will reconsider the hybrid Sorites (SSD) that could not be given a uniform solution in Priest's theory because SI was not locally valid. Now that it is, we can easily come up with such a solution. The recipe is essentially the same as before. Just rerun the line of thought in Chapters 2 and 3, where the two kinds of Sorites (SS and SD) were set up and solved in Łukasiewicz logic (with fuzzy identity). We can rely on the same intuitions here, now applying the new logic $Ł_Z$ drafted in Section 4.4. Rather than two analogous solutions, we obtain one general solution that applies to both subparadoxes. From this point of view, there is no reason whatsoever to maintain that the fuzzy approach would only apply to the vagueness of ordinary relations. Identity can be fuzzy just as well as any relation.

Chapter 5

Objections and replies

Many philosophers have taken great pains to show that the idea of vague identity in general (and fuzzy identity in particular) is impossible. Smith (2008b, p. 12) classifies arguments against vague identity into formal and philosophical ones. We will consider one example from each class, as well as Priest's replies to them (Sections 5.1 and 5.2). We will then give two complementary responses to the philosophical argument (Sections 5.3 and 5.4), the latter of which will expand to a general discussion on the proper notion of identity.

5.1 Evans's argument

The best-known formal argument against vague identity is the one given by Evans (1978),[1] modelled after Barcan–Kripke's proof against contingent identity.

Let a, b be singular terms, and ∇ the sentential operator "it is indeterminate whether." In order to abstract a property from a formula $A(x)$, we will introduce the notation $\lambda x[A(x)]$ such that the predication $\lambda x[A(x)]y$ is equivalent to $A(y)$. We start from the assumption that it is indeterminate whether a is identical to b,[2]

$$\nabla a = b \tag{5.1}$$

from which it follows that b has the property of "being indeterminately identical to a,"

$$\lambda x[\nabla a = x]b \tag{5.2}$$

---

[1] See also Salmon, 1981, pp. 243–245.

[2] For our purposes, we could as well read the operator as "it is vague whether." Even if Evans is here talking about indeterminacy instead of vagueness, what he means by the former is a result of the latter (see the first paragraph of his paper; cf. also p. 5, n. 1).

On the other hand, from the axiom that a is determinately identical to itself,

$$\neg\nabla a = a \tag{5.3}$$

we conclude that a does not have the property of "being indeterminately identical to a,"

$$\neg\lambda x[\nabla a = x]a \tag{5.4}$$

We have thus found a property which b has but a has not. Now, according to Leibniz's Law, two identical objects share all the same properties, the contraposition of which is that if two objects differ with respect to some property, then they are not identical. We therefore infer that the identity statement $a = b$ is false,

$$\neg a = b \tag{5.5}$$

which apparently contradicts our original assumption (5.1) that the statement is of indeterminate truth value.[3]

Evans's original presentation of the argument is notoriously brief and insufficiently explained. As Smith (2008b, p. 13) points out, the precise morale of the argument depends on the semantics for the propositions involved. Nothing stops proponents of vague identity from coming up with model theories that either make the argument invalid or the conclusion (5.5) compatible with the premiss (5.1), to the effect that the argument is not a genuine *reductio* after all. Priest's theory of fuzzy identity is an example of the former strategy. A simple counterexample suffices to refute the argument. Suppose $d(a, b) = 0.3$, and

$$\nabla x = \begin{cases} 1, & \text{for } 0 < x < 1 \\ 0, & \text{for } x = 0 \text{ or } x = 1 \end{cases}$$

It follows that $[\nabla a = b] = 1$ and $[\neg\nabla a = a] = 1$ but $[\neg a = b] = 0.3$. The argument is not valid, not even locally: if ε is set to e.g. 0.6, then the premisses are acceptable for all δ and yet the conclusion is not acceptable, because the truth value of the conclusion does not depend on that of the premisses in any way. The fallacy is due to applying Leibniz's Law with the non-smooth predicate $\lambda x[\nabla a = x]$. It is smooth only if $|f_{\lambda x[\nabla a = x]}(a) - f_{\lambda x[\nabla a = x]}(b)| \leq d(a, b)$, which is clearly not the case (recall the definition on p. 30).[4]

---

[3] Evans actually suggests that 5.5 could be strengthened to yield a straightforward contradiction with 5.1, but we do not have to go into that.

[4] See Priest, 1998, p. 336; Priest, 2008, pp. 577–578.

We might add that the argument is successful only if $\lambda x[\nabla a = x]$ is counted among the generating properties of =, and we can plausibly propose principled reasons for thinking that it is not. To begin with, $\lambda x[\nabla a = x]$ is explicitly defined in reference to =, which has a hint of circularity about it. Predications of $\lambda x[\nabla a = x]$ are not atomic formulae and do not fit in the kind of logic suggested in Section 4.4. Only if we choose to overlook the syntactic appearance of $\lambda x[\nabla a = x]$ and treat it as a primitive property to be counted among the generators does the argument go through, but not even then does it really refute the theory of fuzzy identity: it merely demonstrates that classical identity is a special case of fuzzy identity, as it should be.

I think it is fair to say that purely formal arguments like Evans's are not enough to settle the debate over vague identity. As such, a formal argument only shows that some conclusion can be derived from some premisses by some rules of inference. No *philosophical* judgement can be extracted from it unless an interpretation is given in which the premisses are true, the inference is valid, and the meaning of the conclusion is explained. Following Smith's classification, we could say that formal arguments need to be supported by philosophical arguments in order to succeed.

5.2 Smith's argument

Smith's own argument is a philosophical one, and one that he claims to be more fundamental than any pre-existing argument. In a nutshell, he argues that in order to make sense of a phenomenon, we must model it set-theoretically; but we cannot model vague identity set-theoretically; therefore, we cannot make sense of vague identity (Smith, 2008b, p. 8).

The idea behind the first premiss is to provide a minimum requirement for making clear sense of a phenomenon. If something can be modelled in set theory, then at least we know that it is logically consistent, if nothing else (Smith, 2008b, pp. 2–3). This premiss is in line with the framework of this thesis, which is why we will accept it without further ado and focus on the second premiss. Smith defends it by the privileged position that identity occupies in every set-theoretic model. Unlike any other relation, identity is so basic to our understanding that it emerges directly from the definition of a set: "as soon as we have a set of objects, we have the identity relation thereon" (Smith, 2008b, p. 5). It is the relation in which every object stands to itself and nothing else, and that relation is perfectly precise. We can of course avoid making use of that relation and interpret the identity predicate as a relation other than identity, but that relation is not the 'real' identity, for real identity is already there, built into every model as a "factory-installed" relation that cannot be removed.

Smith takes up examples of theories that purport to make sense of vague identity. He cites

Trillas and Valverde's (1984) view that "indistinguishability . . . overlays identity—it does not supplant it" (p. 232) as an acceptable way to understand the status of identity-like relations (Smith, 2008b, pp. 7–8), whereas Priest's (1998) theory of fuzzy identity is given as an example of a failed attempt to replace 'real' identity (Smith, 2008b, pp. 5–6). Priest has interpreted the identity predicate as a fuzzy relation, defined by the distance metric on the domain, which is not the real identity relation as far as Smith is concerned. Priest has anticipated this objection and replied that the objector is begging the question:

> Plausibly, the fuzzy truth conditions do not define an *ersatz* identity relation, but the genuine thing for vague objects. . . . The objector does not even have a right to *assume* that the domain D is furnished with a relation of the kind required for classical identity. (Note that the semantics given make no use of identity as applied to objects in D. The metalanguage contains a crisp identity predicate, but it is only ever applied to truth values.) (Priest, 1998, p. 337)

In order to make sense of this disagreement, we have to distinguish between two equivalent formulations of the first clause in the definition of a distance metric (Definition 3.1.2). The first formulation, $d(x, x) = 0$, makes no explicit reference to Smith's real identity. There is an occurrence of the symbol '=' but it refers to the identity relation on the set of truth values $[0, 1]$, not the identity relation on the domain D. The second formulation is $d(x, y) = 0$ *if* $x = y$, where the latter occurrence of '=' refers to Smith's real identity. Priest (1998) has conveniently chosen the first formulation over the second, which apparently is the reason for him to deny that the domain be furnished with classical identity. However, if we reason as Smith does (albeit in a context where he is not directly addressing Priest's theory), the first formulation does make an *implicit* reference to classical identity, because the two occurrences of x in $d(x, x)$ are to denote one and the same object (Smith, 2008b, pp. 3–4).

Smith also points out that Priest has excluded the converse clause, $d(x, y) = 0$ *only if* $x = y$, which is often included in the definition of a distance metric. "Whilst this condition could be added here also, it plays no significant part in what follows," Priest writes (1998, n. 5), ignoring the fact that its inclusion would actually have forced him to make explicit reference to classical identity (Smith, 2008b, n. 5).[5] The central question, then, is whether Priest is right in thinking that classical identity can be avoided by avoiding any explicit reference to it in the model. Not only does Smith insist that Priest fails to do away with classical identity, but he also seems to imply that Priest's definition of identity is actually *dependent on* classical identity, regardless of

---

[5] In the later incarnation of his theory, Priest (2008, § 25.6) has in fact replaced the first clause with "$d(x, y) = 0$ iff x is y," now including the fourth clause that had earlier been excluded.

Priest's claim to the contrary.

I think there is a sense in which Smith is right. Recall how we constructed our original model for the fuzzy identity of Theseus's ship (pp. 29–31). We presupposed a domain of objects D, defined a distance metric d on D, and then defined the interpretation of the identity predicate as a fuzzy relation on D by means of d. The domain D is therefore conceptually prior to fuzzy identity. Let us forget about fuzzy identity for now and concentrate on the domain, which is nothing more than an ordinary (crisp) set of elements $s_0, s_1, \ldots, s_{1,000}$. As we know from any introductory course on mathematics, a set is defined by its elements: two sets are identical if and only if they have the same elements. From this point of view, each $s_n \in D$ is a distinct element for the simple reason that we defined D so. When we defined the set D as $\{s_0, s_1, \ldots, s_{1,000}\}$, we implicitly defined a primitive relation such that each s_n is identical to itself and nothing else, which is the 'real' identity on D.

Be that as it may, I do not think Priest is completely wrong, either. Even if Smith is right to assume the presence of real identity, the identity he is talking about is not on the same level as Priest's fuzzy identity. As I see it, Priest's fuzzy identity is *defined* by means of Smith's real identity, which is in turn a *presupposed part of* the metalanguage. Priest may be mistaken in denying the dependence of fuzzy identity on real identity, and the presence of real identity in the model altogether, but his mistake is understandable given that real identity is not defined in the way relations usually are (cf. Smith, 2008b, pp. 4–5).

Smith concludes his paper by developing a point raised by Williamson (2002, pp. 283–284) on the relationship between a vague object language and the metalanguage it is modelled in. If the metalanguage is classical (as it apparently is in Priest), vague identity is interpreted either as real identity, which is not *vague*, or as another relation, which is not real *identity*. That is one way to sum up everything we have discussed so far. What we have not considered yet, however, is the possibility that the metalanguage itself be vague. Smith's rebuttal of such a possibility is based on the first premiss of his main argument. In order to make sense of vague identity, we need to present a model theory using the resources of standard set theory; but if the model theory is given in a metalanguage which is considered vague in the same sense as the object language for which it is given, we are right back at where we started from: we do not understand the vagueness of the metalanguage unless it is explained to us in standard set theory, or some alternative way which has not yet been presented and probably never will (Smith, 2008b, pp. 13–15).

In the later incarnation of his theory, Priest (2008, p. 578) has changed his attitude toward the objection that his fuzzy identity is not really identity.[6] He acknowledges that the identity

---

[6] Whether Smith had any direct influence on Priest's change of attitude, I do not know; however, see

relation of the semantics is the standard crisp one, and he seriously considers the possibility that the objects of the theory are therefore crisp, after all, dedicating a separate "methodological coda" to discussing this objection and the more general issues it raises (Priest, 2008, pp. 584–586). If the objection is taken seriously, the identity of the object language "is not really identity, just some sort of similarity relation."[7] Priest continues:

> There is something wrong about this objection, and something right. It is certainly the case that the identity relation of the object language and the identity relation of the metalanguage . . . are different. It does not follow that it is the relation of the object language that is not the real notion. (Priest, 2008, p. 584)

Priest admits that it would be better if the semantics were given using the notions of the object language, or if the notions of the metalanguage could at least be "made sense of" using the notions of the object language so that the semantics thus specified "makes perfectly good sense." Although it is not a general requirement in logic that the semantics of a language should always be given in terms of the same notions as those of the language itself—the interpretation of modal operators in terms of possible worlds is a good example—it is still a failing for many non-classical logics if their semantics rely on the very classical notions that they are meant to replace (Priest, 2008, pp. 584–585). In some cases, it is possible to specify the semantics in a way that avoids any problematic dependence on classical logic,[8] but in the case of fuzzy identity, such a specification has not been given (Priest, 2008, pp. 585–586). All in all, it seems as if Priest is ready to admit that the reliance of fuzzy identity on crisp identity is a real problem for his theory, but he does not even attempt to solve it.

In what follows, I will give two complementary responses to Smith's objection against fuzzy identity. The first one continues the previous discussion about modelling fuzzy identity in a precise metalanguage and denies that this be a problem in itself (Section 5.3). The other response goes to a more fundamental level, questioning Smith's understanding of identity altogether (Section 5.4).

Smith, 2008b, n. 1.

[7] This comment also raises a terminological issue about the proper meanings of the terms 'identity' and 'similarity relation.' In Section 3.1, we used the latter term to name a notion which was then in Section 3.2 used to define Priest's notion of fuzzy identity. Priest (1998) does not use that term at all, whereas Priest (2008, pp. 580–581) explains that fuzzy identity could be defined in terms of a 'similarity function' as well. We will touch upon this terminological issue in Section 5.4.

[8] Priest's examples of such cases are the intuitionistically acceptable semantics of intuitionistic logic in Dummett (1977, Ch. 5), and the paraconsistently acceptable semantics of paraconsistent logic in Priest (2006, Ch. 18).

5.3 Methodological interlude

My first response to Smith continues immediately from where the previous section ended. I agree that Priest's theory bases fuzzy identity on crisp identity, but I challenge the assumption that it would be a failure to do so. On the contrary, I argue that reliance on crisp semantics is an intentional methodological choice that lies at the very heart of the fuzzy approach.

Priest does admit that "the identity relation of the object language and the metalanguage are out of kilter" and that there is "something *prima facie* awry in the situation" (2008, p. 584). But identity is not the only thing out of kilter. All fuzzy notions in the object language depend on crisp notions in the metalanguage one way or another. Think of the fuzzy membership predicate '$\in$' of Example 4.4.4, for instance. Just as it can be objected that the interpretation of '$=$' as $Z_Ł$ is not the real identity relation, it can be objected that the interpretation of '$\in$' as $f_\in$ is not the real membership relation. The metalanguage already has a membership relation of its own, and that relation is totally crisp. From the definition of the domain D of $\mathcal{M}$, it follows e.g. that $a^{\mathcal{M}} \in D$ and $q^{\mathcal{M}} \in D$. Whether $a^{\mathcal{M}} \in q^{\mathcal{M}}$ needs not be determined in the model—in this respect, the situation is admittedly different from identity—but if it were determined, it would be necessarily precise. To say that it is vague in the same manner as the formula $a \in q$ is vague in the object language only takes us back to square one. A similar argument could plausibly take down fuzzy inclusion (the relation "is a subset of"), fuzzy ordering (Zadeh, 1971), fuzzy congruence, fuzzy divisibility, or any other attempt to fuzzify a notion of standard mathematics. Every such an attempt is bound to build on standard crisp semantics if it is to succeed.

Of course, this is not to say that *all* relations be necessarily crisp in the same sense as identity. Natural language predicates such as "heap" are not parts of the metalanguage, and so it might be tempting to think that they could still be fuzzy even if identity could not. But even their fuzziness is modelled in the metalanguage relying ultimately on crisp notions. Suppose the predicate H is interpreted as a function f_H on $D = \{a, b, c, d, e\}$ such that

$$f_H(a) = 0 \qquad f_H(b) = 0.25 \qquad f_H(c) = 0.5 \qquad f_H(d) = 0.75 \qquad f_H(e) = 1$$

We tend to think of $f_H \colon D \to [0, 1]$ as the characteristic function of a fuzzy unary relation. Yet the function f_H itself is a crisp function; what is more, it can be reduced to a crisp binary relation $R_H \subseteq D \times [0, 1]$,

$$R_H = \{(a, 0), (b, 0.25), (c, 0.5), (d, 0.75), (e, 1)\}$$

Any fuzzy relation of arity n can be similarly reduced to a crisp relation of arity $n + 1$. This is strong evidence for the conclusion that crisp notions are conceptually more fundamental than fuzzy ones. In order to make sense of fuzzy notions, we must rely on crisp ones.

In light of the preceding remarks, it seems as if the impossibility to replace crisp identity in the metalanguage is just the top of the iceberg: the extreme example of a general phenomenon which is characteristic of the fuzzy approach as a whole. The discrepancy between a vague object language and a precise metalanguage is a design choice, accepted as a fundamental assumption of the paradigm. This assumption was explicitly stated already in J. A. Goguen's seminal paper that laid the foundation for the fuzzy approach to vagueness:

> Our models are typical purely exact constructions, and we use ordinary exact logic and set theory freely in their development. This amounts to assuming we can have at least certain kinds of exact knowledge of inexact concepts. ... It is hard to see how we can study our subject at all rigorously without such assumptions. (Goguen, 1968–1969, p. 327)

One can of course maintain that the preciseness of the metalanguage is a problem, but it is then a problem for the entire fuzzy approach and not just for fuzzy identity. If one wants to make an exception for identity, it is not sufficient to point out that it is modelled in a precise metalanguage: one must seek justification from elsewhere, such as the special position of identity in philosophy.

5.4 How real is Smith's 'real identity'?

Smith has argued that what he calls 'real identity' cannot be vague. Suppose we are now convinced that he is right. Are we to conclude that identity cannot be vague? As so often in philosophy, the answer depends on what we mean by 'identity'. Even if Smith's 'real identity' cannot be vague, we may still ask whether it is indeed *real identity*. Smith is right only insofar as identity really is the kind of relation that he takes it to be: i.e., the identity of individuals that is presupposed by the definition of a set.

My second and more fundamental response to Smith's argument is to question his notion of 'real' identity and argue for a broader understanding of identity in philosophy. While I agree that Smith's 'real' identity cannot be vague, I suggest that identity in general could be. My response has three parts: first, I demonstrate that Smith's identity is a trivial notion that fails to account for any of the philosophical puzzlement about identity (Section 5.4.1); second, I will come up with a classification of different conceptions of identity and point out that on that classification, Smith's identity is an extreme conception (Section 5.4.2); and third, I will argue that vagueness is logically possible for virtually any conception of identity on the given classification apart from Smith's (Section 5.4.3).

While my response does not actually *refute* Smith's argument—to do that would require much more than what is possible within these pages—it will anyway give us strong reasons to regard his argument with suspicion. It will show that Smith's identity cannot be vague just because he defines identity so that it cannot be vague. The conclusion of his argument is true by definition, which makes it seem as if he is not as much giving an argument as stating an obvious fact. The real question is not whether his definition of identity allows vagueness, but whether his definition of identity is *adequate*. Smith does not show that it is, and I will give strong reasons to assume that it is not.

5.4.1 Trivial identity

Smith's 'real' identity is a purely formal and primitive relation that cannot be further analyzed or explained in any way. It is sufficient to have a set of objects and there it is, identity: a relation that automatically holds between any object and itself and nothing else. We may of course ask what the objects in the set really are or what properties they possess, but none of that is essential to Smith's conception of identity. I aim to argue that this conception is trivial and fails to account for any of the philosophical puzzlement about identity.

A concrete example should make this clear. Let us define a set $D = \{a, b\}$. According to Smith, we now have an identity relation on D. We then ask whether that relation holds between a and b. Is it the case that $a = b$? It all depends on how we defined the set D in the first place. If by saying $D = \{a, b\}$ we meant (as we usually do) that a and b are distinct elements, then of course they are not identical. If we meant that a and b are one and the same object that we chose to list with two different symbols for some unthinkable reason, then of course a and b are identical. Whether they are identical or not, it is nothing more than a precondition that should have been decided on when we first defined the set. If the elements of the set have not been given clear identity conditions, then the set itself has not clear identity conditions; for then it is unclear whether, e.g., $\{a, b\}$ is the same set as $\{a\}$. Once we have defined a set, we have defined the identity conditions for its elements, or otherwise we have not really defined the set. From this point of view, to say that the definition of D presupposes the identity of its elements is almost as self-evident as saying that it presupposes the relation "is an element of D." Identity in this sense is so self-evident that it is hard to say anything about it without it being circular, trivial or utterly uninformative—so self-evident that it is hard to see why we even have to say it—and yet that is all there is to say about identity, if we define identity as Smith does.

In the beginning of his paper, Smith (2008b, pp. 1–2) takes up three alleged examples of vague identity (one of which is Theseus's ship) and promises to try to explain why we cannot make sense of vague identity in a way that sheds light on why we have found it so puzzling. I daresay he fails to achieve this aim. At no stage in his paper does he return to these examples or explain what we should make of them in light of his conception of identity. Not only do his views exclude vague identity, but they also exclude all non-trivial cases of classical identity. Take the identity $s_0 = s_1$, for instance. I believe most people would accept it as a genuine case of classical identity, or at least it is in accord with common sense that a ship can persist the replacement of one plank. However, if we define identity as Smith does, then $s_0 = s_1$ is false just because s_0 and s_1 are interpreted in the model as two distinct elements of the domain—or, in an alternative model, true just because s_0 and s_1 are interpreted as one and the same element of the domain. Either way, the real problem about whether a ship can persist the replacement of one plank has been solved already before logic steps in. For Smith, that is not a problem about identity. Indeed, he does not even discuss the philosophical business that has to be settled before one can come up with a crisp model where the identity $s_0 = s_1$ either holds or not. That is to disregard that there be a genuine problem about identity, which does no justice at all to the intuitions at the heart of the puzzle.

Identity in Smith's sense can tell us nothing about real objects. Rather than representing any state-of-affairs in the world, it only reflects a precondition about the formal *apparatus* representing the world. If there is any problem about $s_0 = s_1$ on Smith's account, it must be either because we do not know whether the interpretation function maps s_0 and s_1 to the same object in the domain, or because the function has not been determined for s_0 and s_1. As soon as the denotations of s_0 and s_1 have been determined and we know them, there cannot be any problem about whether they are identical. Once it has been decided which term denotes which object, every identity statement automatically comes out as either true or false.

It remains to be seen whether there is more to identity than Smith's very narrow understanding of it. The plentiful discussion among philosophers on vague identity, not to mention identity in general, could be considered evidence for a yes. When metaphysicians talk of identity, they are interested in questions of the following kinds:

- Is Tibbles identical to Tib? (See Wiggins, 1968.)
- Is Goliath identical to Lumpl? (See Gibbard, 1975.)
- Is the original Theseus's ship s_0 identical to $s_{1,000}$?
- Am I the same person as I was yesterday?
- Is the actual Trump identical to his counterpart in a counterfactual world?

None of these puzzles are solved by logic alone. The real puzzlement is elsewhere than a logical language or its model theory: they are nothing more than technical apparatus to formulate and evaluate statements about the world. One can always come up with a model and show that the identity in question holds or does not hold in that model, but that would be begging the question. The question is not about whether some identity holds in some model; it is about which model is metaphysically correct.

Be that as it may, Smith may very well be content with replying that none of the aforementioned problems are about 'real' identity, and that is the end of the story. He is not alone with his views; on the contrary, the view of identity as a trivial notion has a long tradition in philosophy. It can be traced back in time to as far as Heraclitus, who famously argued that one cannot bathe twice in the same river. Similar positions are found in e.g. Hume (1739–1740/2010), Quine (see Béziau, 2003), and Lewis (1986, Ch. 4), who argues that there are no philosophical problems about identity. He does admit that the problems are real problems but points out that they can be rephrased without any mention of identity, which he takes to show that they are not about identity after all. Identity as such

> is utterly simple and unproblematic. Everything is identical to itself; nothing is ever identical to anything else except itself. There is never any problem about what makes something identical to itself; nothing can ever fail to be. And there is never any problem about what makes two things identical; two things never can be identical. (Lewis, 1986, pp. 192–193)

Lewis goes on to propose a weaker relation, *counterparthood*, which is to replace identity across possible worlds. Identity across time is likewise supplanted by Lewis's four-dimensionalism, i.e., the view that identifies a persisting object with a compound of temporal parts. On this view, none of the three-dimensional objects $s_0, \ldots, s_{1,000}$ are identical with each other nor the 'real' Theseus's ship, which is a four-dimensional object. The relation of $s_0, \ldots, s_{1,000}$ to the ship is not identity but *parthood*.

Hume, Quine, Lewis and Smith all have a very narrow conception of identity. As identity as such is unproblematic, what is problematic about identity is something else than identity. This conception is aptly formalized by the standard theory of identity in first-order logic. Priest (2010, p. 406) agrees that the standard theory is unproblematic, but only because it is formulated precisely so as to exclude everything that is problematic about identity. He goes on to discuss several puzzles about identity (including a variant of Theseus's ship, again), and concludes that the properties classically attributed to identity

> are not to be taken for granted philosophically. One can, of course, simply specify *by*

> *fiat* that identity has these properties. . . —and call it identity if you like; but it is all too obvious that the behaviour of the relationship involved in the above examples—and which we used to call identity before the word was usurped—still cries out to be understood. (Priest, 2010, p. 407)

As noted, Smith does nothing to explain the behavior of the relationship involved in the puzzles that many people think to be about identity. He has not explained why they would not be problems about identity or given any reasons why the notion of identity should be restricted to his 'real' identity. His main point is that if we dig deep enough into the architecture of a set-theoretic model, we will always find a precise identity relation presupposed by the domain, but he does not say why this relation alone should be called identity.

5.4.2 From total identity to partial identity

If Smith's conception of identity is trivial, what then would be an adequate conception of it? This question is too big to be given a conclusive answer here; it is just the kind of a metaphysical question that was deliberately excluded from this thesis. Instead of arguing for a specific conception of identity, we will provide a classification of logically available philosophical conceptions of identity, where Smith's conception of identity turns out to be an extremist one. Such a classification is of course not a refutation of Smith's argument, but it will anyway pose a challenge to Smith, who should be able to explain why all the other conceptions fail.

Some terminological distinctions are in order. The first step is to distinguish *numerical identity* from *qualitative identity*. Two things are qualitatively identical if they have properties in common, and numerically identical if they are in fact one and the same thing (Noonan & Curtis, 2018, § 1). For instance, a banana is qualitatively identical to a lemon with respect to color, and yet numerically distinct from it, but it certainly is numerically identical to itself. Whereas numerical identity is something that either holds or not, qualitative identity is more like a matter of degree. Not only may some two things be more or less qualitatively identical than some other two things (e.g., a banana and a lemon may be more qualitatively identical than a banana and a turnip), but also the same two things may be more or less qualitatively identical depending on which properties are considered relevant (e.g., a banana and a lemon are more qualitatively identical with respect to their color than e.g. their shape).

The relationship between numerical and qualitative identity is not entirely clear. Most importantly, it is controversial whether total qualitative identity is a necessary condition for numerical identity or whether it is actually a sufficient one. The difficulty is akin to the undefinability of identity in first-order logic (p. 15) and the controversy surrounding the identity of indiscernibles (see e.g. Black, 1952). Everett W. Hall has suggested another distinction that should help us make sense of not only numerical and qualitative identity, but different conceptions of identity more generally:

> The controversy between those who claim that there is such a thing as merely numerical identity and those who assert that all identity is qualitative is clarified, it seems to me, if we view it as concerned essentially with the following question: What difference of properties or characteristics is incompatible with the identity of a concrete, existent thing? There are two extreme answers to this question. It may be said that any difference whatever is incompatible with identity in any instance. Again, it may be said that no difference whatever is incompatible with any identity. (Hall, 1933, p. 88)

We will call the kinds of identity suggested by the two extreme answers *total* (or *absolute*) *identity* and *universal identity*, respectively.[9] As I see it, total identity holds between any two individuals if and only if they are exactly alike in *all* properties and can only hold between an object and itself, whereas universal identity does not require sameness of any properties and can hold between any two objects, no matter how different. Between the two extremes, there are as many kinds of identity as there are sets of differentiating properties. The more properties are included, the more fine-grained is the partition of objects so produced, and the closer the resulting relation is to total identity. Each such identity obeys the laws of LL and II with respect to the properties in question. We might call these intermediate kinds of identity *partial identity*.[10]

Hall goes on to note that proponents of numerical identity have usually rejected total identity, as it was defined above, in favor of a more moderate conception of identity as being incompatible with only one kind of difference, namely, a numerical (as opposed to qualitative) one. Hall does not know of anyone who has embraced total identity in its extreme form, but he thinks that it has at least been suggested by Hume and Russell (Hall, 1933, pp. 88–89). The other extreme has, according to Hall, been supported by absolute idealists such as Bradley and Bosanquet (Hall, 1933, pp. 101–102). Intermediate conceptions of identity, based on a specific property or class of properties, are attributed by Hall to Plato, Aristotle, McTaggart and Locke, among others.

It seems that the doctrine of absolute identity has gained more popularity since the publication of Hall's paper in 1933. In light of our previous remarks (in Section 5.4.1), we should add at least Lewis and Smith to the proponents of absolute identity. But it has also been famously

---

[9] This is an intentional deviation from Hall, who has chosen to call the latter kind 'partial identity.' I think it is more appropriate to reserve that title for the relations between the two extremes.

[10] Alternatively, we might think of the different kinds of identity as the different senses of the word 'identity.' On this view, total or absolute identity is identity in the strict sense, to be distinguished from identity in the loose sense (cf. Butler, 1736/1849; Chisholm, 1969; Baxter, 1988).

objected by Geach (1967), building on Locke (1690/2004, Ch. XXVII). Geach argues that all identity is *relative*: the question is not whether e.g. s_0 is identical to $s_{1,000}$ *simpliciter*, but rather whether s_0 is the same *ship* as $s_{1,000}$. If he is right, it follows that Smith's 'real' identity is either not real (because there is no absolute identity), or then it is really relative identity and so it has no exclusive right to the title 'identity' (see n. 11 below).

Luckily, the framework of partial identity can accommodate Geach's relative identity as well; as a matter of fact, Hall already had similar ideas himself (see Hall, 1933, pp. 103–104). Relative identity can be made sense of in terms of partial identity, for every partial identity is relative to its differentiating properties. One might suggest, for instance, that s_0 and s_1 are the same *ship* because they are alike in all such properties that are relevant to "shiphood." On the other hand, they are surely not the same *temporal part*, as they are distinguished by the properties of "existing at the time t_0," possessed by s_0 but not by s_1, and "existing at the time t_1," possessed by s_1 but not by s_0. The identity of ships and the identity of temporal parts are both kinds of relative identity that can be put on the spectrum drafted above.

Much more could have been said on different conceptions of identity, but I hope to have nevertheless convinced the reader that Hall's framework provides us with a conceptual map where all prominent views on identity fall in place; and, as we have seen, Smith's view is thereby placed in the radical left (or right, if you like).

5.4.3 Spectrum of similarity

The distinction between total and partial identity allows us to conclude that while we agree with Smith on the thesis that total identity cannot be vague, we have yet to find out whether partial identity can be vague. It is here that the interpretation of fuzzy identity as similarity steps in. Similarity relations, as they have been defined and analyzed in this thesis, are precisely the logical realization of the different conceptions of identity; and, as we have seen, similarity relations can be vague (fuzzy) if they are generated by vague (fuzzy) properties.

As noted already in Section 4.3, similarity relations can be put on a scale according to their sets of generators. At one end, we have the relation $Z_{[0,1]^D}$ with the maximal set of generators. Every element in D stands in $Z_{[0,1]^D}$ to itself and itself only. This is because for any two distinct elements in D, there is at least one property that sets them apart; specifically, for each $x \in D$ there is the property of "being classically identical with x," distinguishing it from all other elements. Obviously, $Z_{[0,1]^D}$ coincides with the classical identity on the domain D, embodying the idea of 'total' or 'absolute' identity presented above.[11] At the opposite end, we have the

---

[11] We might add, in the spirit of Geach, that even this 'absolute' identity is actually relative to the underlying theory and the language in which it is stated. For instance, s_0 is one primitive element of D

relation $Z_\emptyset$ with the minimal (empty) set of generators. All elements in the domain D stand in $Z_\emptyset$ to each other, for there is no property to distinguish them—this is the logical explication of 'universal identity.' In between $Z_{[0,1]^D}$ and $Z_\emptyset$, there is a spectrum of similarity relations Z_Φ with sets of generators Φ between $[0,1]^D$ and $\emptyset$ (i.e., $[0,1]^D \supseteq \Phi \supseteq \emptyset$). Each similarity relation on the spectrum corresponds to a specific sort of 'partial identity' complying with a Leibniz's Law of a specific strength.[12]

This spectrum of similarity relations is also a spectrum of solutions to the identity Sorites (SD). If identity is interpreted as $Z_{[0,1]^D}$, we end up with a completely fallacious argument from a fully true minor premiss and a fully false major premiss to a fully false conclusion. If identity is interpreted as $Z_\emptyset$, we are stuck with a perfectly sound argument from fully true premisses to a fully true and yet paradoxical conclusion. The latter solution embraces the intuition that created the paradox but does nothing at all to solve it, whereas the former one only technically "solves" the paradox but gives no explanation at all to the intuition that created it in the first place. In between the two extremes, we have a spectrum of solutions, each of which amounts to a specific balance between the two desiderata.

I anticipate that I will be blamed for being too liberal with my use of the word "identity." Although I have made it clear that it is not my intention to argue for any particular conception of identity, it is true that when I speak of "different conceptions of identity," I am already making a preliminary judgement about which relations *possibly* count as identity. It can be complained that the class of partial identity, as I have defined it, is all too broad: the smaller the set of generators, the less does the resulting relation have anything to do with identity. It may seem as if the notion of identity undergoes inflation unless we draw a line somewhere. But it should be clear by now that the fuzzy theorist is very reluctant to draw any sharp lines. Plausibly, if we were to narrow down the possible conceptions of identity, they would comprise a subspectrum from total identity up to some point, but the exact location of that point is controversial because the conflicting desiderata of formal rigor and philosophical applicability are pulling it to opposite directions. How far from total identity can a relation be and still be identity in the loose sense? Can a balance be found between the rival requirements?

There is a more fundamental objection that is relevant regardless of whether a balance can be found in the end; an objection that brings us back to the central point of Smith's argument against vague identity. Despite our efforts to make room for fuzzy identity, it could still be argued that fuzzy identity is not identity even in the loose sense; rather, it is a fuzzy version of

---

in our metalanguage, but we can conceive of a "metametalanguage" where s_0 is further analyzed as an equivalence class of several contemporaneous objects.

[12] The ends of the spectrum correspond to two of the four relations named by Tarski as "the only logical binary relations between individuals," whereas the other two are their negations (1986, p. 150).

counterparthood, the weaker alternative to identity that has been suggested to replace not only identity across possible worlds (Lewis, 1986) but also identity across time (Hawley, 2004; Sider, 2001). On this view, it is a mistake to call similarity relations by the name "identity." Fuzzy "identity" amounts to similarity, not identity.

I am open to this objection. I acknowledge that my use of the word "identity" is provocative in this context. To many people, it appears to be a sacred word that they want to reserve for 'real' identity only. If that is so, the debate reduces to an unphilosophical pedantry about which meaning is appropriate for the word in question, but that is perfectly okay; let them keep that word if they like. My thesis does not require a crisp distinction between identity and counterparthood, anyway. On the contrary, fuzzy logic is employed precisely to allow for degrees of similarity. Instead of a dichotomy of two rival relations we have a spectrum, one end of which is total similarity or sameness, i.e., 'real' identity. This allows us to examine different definitions of similarity and see how close they are to real identity. Fuzzy identity will in fact bridge the gap between identity and counterparthood and pave the way towards a compromise between the two approaches.

Another way to make sense of this is to say that while identity in the strict sense is precise, the notion of identity in the loose sense is vague. Not only can a given identity relation be first-order vague in the sense that it accepts fuzzy truth values, but there is also second-order vagueness as to whether a given similarity relation counts as identity. We might even make this to a second-order Sorites paradox. Let $Z_0, \dots Z_m$ be a sequence of similarity relations generated by $\Phi_0, \dots, \Phi_m$, respectively, such that $\Phi_0 = [0,1]^D$, $\Phi_m = \emptyset$ and $\Phi_n \supset \Phi_{n+1}$ for all $n = 0, \dots, m-1$. The reader will probably see where this is going:

$$\begin{array}{l} Z_0 \text{ is identity} \\ \text{If } Z_n \text{ is identity, so is } Z_{n+1} \ (n = 0, \dots, m-1) \\ \hline Z_m \text{ is identity} \end{array}$$

Of course, the default solution on the fuzzy approach is to interpret "Z_n is identity" as a fuzzy predication, the truth value of which depends on the position of Z_n on the spectrum.

One could object that the second-order paradox gives instantly rise to a third-order paradox about its correct solution, and so on, *ad infinitum*. That is a fair point, but identity is no different from other relations in this respect. My point is just to argue that identity be treated equally with other relations. To draw a sharp line between (strict) identity and similarity would be against the very intuitions that motivate the fuzzy approach in the first place. For a fuzzy theorist, it might in fact be more natural to think of similarity as the overarching concept of interest and identity as an uninteresting limiting case of it.

Conclusion

Priest concludes his original presentation of the theory of fuzzy identity as follows:

> We have seen that a semantics can be given for fuzzy identity which is natural, and which also brings out the similarity between the two different kinds of sorites in a natural way. It makes possible the notion of *local validity*, which provides for a simple and plausible diagnosis of the mistake in sorites arguments. We have also seen that the semantics shows the invalidity of [Evans's] objection to fuzzy identity. As a theory of fuzzy identity, little more could be asked of it. (Priest, 1998, p. 340)

In Section 3.4, we saw that there is actually quite a lot to be asked of it, and we have consequently developed the theory so as to meet those expectations. We have shed more light on the nature of the distance metric and smooth predicates, paving the way for a metaphysical foundation of fuzzy identity. We have strengthened Priest's logic with fuzzy identity in a way that makes the rule of SI locally valid without restrictions; the same is true of the rule of LL. The corresponding schemes SI&$^{\mathrm{w}}$ and LL&$^{\mathrm{w}}$ are as valid as Ref, Sym and Tr&. While Priest could only solve the special identity Sorites with transitivity (SD), we have been able to solve the general case with SI (SSD). The standard Sorites and the identity Sorites are thereby merged into one paradox with a uniform solution. The refutation of Evans's counterargument still applies with minor additions. We have also further developed Priest's defense against Smith's counterargument. A continuum from any particular fuzzy identity to classical identity has been established, allowing us to introduce relations that are just as close to classical identity as necessary. Considering all these clarifications, elaborations and adjustments that have been made to the theory of fuzzy identity, would it *now* be safe to say that "little more could be asked of it"?

Well, not really. Much more could have been said on the subject than what has been possible within these pages. The theory could be further developed into various directions, and various aspects of it could be investigated in still more detail. Even so, what we have been able to accomplish here is to have answered the research question that was set at the outset (p. 6). We have given strong reasons to conclude that inside the fuzzy framework, there is no reason to discriminate between the fuzziness of identity and other relations. Fuzzy identity can surely be criticized from outside the framework, and its metaphysical foundation can be questioned very well, but the grounds on which these challenges are based do not arise from fuzzy logic itself. The disagreement on the fuzziness of identity between the two fuzzy theoreticians, Priest and Smith, has thus been settled in favor of Priest.

As for the further developments, one would be to broaden the perspective to cover the general discussion on vagueness. The standard objections to the fuzzy approach or the degree-theoretic approach in general (see Smith, 2008a, Pt. III) could be reconsidered in the case of fuzzy identity. In this connection, it could also be asked whether the introduction of fuzzy identity brings with it elements from other theories of vagueness, such as plurivaluationism (Smith, 2008a, Ch. 6) or epistemicism (MacFarlane, 2010).

Another development would be to expand the syntactic analysis of the Sorites paradox to a full-fledged proof-theoretical treatment of fuzzy identity. The first step would be to finalize the axiomatization suggested in Section 4.4, but the next step would definitely be to come up with a sequent calculus for the logic of fuzzy identity (cf. Section 3.3). To obtain a system with as good structural properties as possible, I would attempt to formulate the axioms as rules applying the method described in Negri and von Plato (2001, 2011). A related, more application-oriented approach would be to consider fuzzy identity in relation to 'fuzzy similarity-based reasoning,' which has been studied by means of such notions as approximate entailment and approximate proof (see e.g. Klawonn & Castro, 1995, § 4; Godo & Rodríguez, 2008).

Finally, it would be worthwhile to take on the metaphysical approach that was acknowledged as the motivating force of this thesis and yet deliberately excluded from it. It would be interesting to come up with a metaphysical interpretation of the fuzzy identity suggested here and see whether it survives known objections to vague identity. Our example case has been about identity across time, but fuzzy identity could be set across space or possible worlds as well. It could be studied in relation to e.g. material composition, contingent identity, essentialism, haecceitism, counterparthood, stage theory, and the realism–nominalism debate, to name just a few possibilities. Forbes' (1985/2016) fuzzy essentialism might be a good place to start.

Specifically, it would be interesting to see what the objects that stand in the relation of fuzzy identity are like. While we have managed to come up with a logic that makes sense of fuzzy identity and subsequently overthrown Evans's argument against vague objects, we have not yet shown (nor was it our intention to show) that there are vague objects; we have only removed a major obstacle in their way. Whether there are vague objects is something that logic alone can never tell us; to find an answer to that question, one must step into the territory of metaphysics.

To those who are unwilling to go there, however, I should make it clear that the application of fuzzy identity far exceeds metaphysics. Poincaré (1905/1958) already points out that outside mathematics, what we encounter between observables is not identity but empirical *indistinguishability*, which, unlike identity, is not transitive: even if x is indistinguishable from y and y from z, it does not follow that x be indistinguishable from z (as cited in Trillas & Valverde, 1984, pp. 253–254). Menger (1951) envisioned that a new kind of relation could be formalized that would restore transitivity. It is against this background that the theory of similarity relations originally arises. The identity Sorites can be formulated for series of sensations, such as a spectrum of indistinguishable hues of a color (Priest, 1998, p. 331). Fuzzy identity applies to this case just as well as the metaphysical one, or even better. Rather than out there in the world, vagueness would then be located in our representation of the world. To find fuzziness, it is not necessary to peek behind the veil of perception, for there is fuzziness all over the veil itself.

Bibliography

Barnes, E. (2010). Ontic vagueness: A guide for the perplexed. *Noûs*, *44*(4), 601–627. https://doi.org/10.1111/j.1468-0068.2010.00762.x

Barnes, E., & Williams, J. R. G. (2011). A theory of metaphysical indeterminacy. In K. Bennett & D. W. Zimmerman (Eds.), *Oxford studies in metaphysics: Volume 6* (pp. 103–148). https://doi.org/10.1093/acprof:oso/9780199603039.003.0003

Baxter, D. L. M. (1988). Identity in the loose and popular sense. *Mind*, *97*(388), 575–582. http://www.jstor.org/stable/2255191

Běhounek, L., Cintula, P., & Hájek, P. (2011). Introduction to mathematical fuzzy logic. In P. Cintula, P. Hájek, & C. Noguera (Eds.), *Handbook of mathematical fuzzy logic: Volume 1* (pp. 1–101). College Publications.

Béziau, J. (2003). Quine on identity. *Principia*, *7*(1–2), 1–15. https://periodicos.ufsc.br/index.php/principia/article/view/14733

Black, M. (1937). Vagueness: An exercise in logical analysis. *Philosophy of Science*, *4*(4), 427–455. http://www.jstor.org/stable/184414

Black, M. (1952). The identity of indiscernibles. *Mind*, *61*(242), 153–164. http://www.jstor.org/stable/2252291

Butler, J. (1849). Of personal identity. *The analogy of religion*. Oxford University Press. (Original work published 1736)

Chisholm, R. M. (1967). Identity through possible worlds: Some questions. *Noûs*, *1*(1), 1–8. https://doi.org/10.2307/2214708

Chisholm, R. M. (1969). The loose and popular and the strict and philosophical senses of identity. In N. S. Care & R. H. Grimm (Eds.), *Perception and personal identity* (pp. 82–106). Press of Case Western Reserve University.

Copeland, B. J. (1997). Vague identity and fuzzy logic. *The Journal of Philosophy*, *94*(10), 514–534. https://doi.org/10.2307/2564549

Cowles, D. W., & White, M. J. (1991). Vague objects for those who want them. *Philosophical Studies*, *63*(2), 203–216. http://www.jstor.org/stable/4320230

Dummett, M. (1959). A propositional calculus with denumerable matrix. *The Journal of Symbolic Logic*, *24*(2), 97–106. http://www.jstor.org/stable/2964753

Dummett, M. (1975). Wang's paradox. *Synthese*, *30*(3/4), 301–324. http://www.jstor.org/stable/20115034

Dummett, M. (1977). *Elements of intuitionism*. Oxford University Press.

Edgington, D. (1992). Validity, uncertainty and vagueness. *Analysis*, *52*(4), 193–204. https://doi.org/10.2307/3328335

Edgington, D. (2000). Indeterminacy de re. *Philosophical Topics*, *28*(1), 27–44. http://www.jstor.org/stable/43154330

Evans, G. (1978). Can there be vague objects? *Analysis*, *38*(4), 208. https://doi.org/10.1093/analys/38.4.208

Fine, K. (1975). Vagueness, truth and logic. *Synthese*, *30*(3/4), 265–300. http://www.jstor.org/stable/20115033

Forbes, G. (1983). Thisness and vagueness. *Synthese*, *54*(2), 235–259. http://www.jstor.org/stable/20115837

Forbes, G. (2010). Identity and the facts of the matter. In R. Dietz & S. Moruzzi (Eds.), *Cuts and clouds: Vagueness, its nature, and its logic* (pp. 419–437). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199570386.003.0025

Forbes, G. (2016). *The metaphysics of modality* (Web Edition). https://spot.colorado.edu/~forbesg/metmodpage.html (Original work published 1985)

Geach, P. T. (1967). Identity. *The Review of Metaphysics*, *21*(1), 3–12. http://www.jstor.org/stable/20124493

Gentzen, G. (1969). Investigations into logical deduction. In M. E. Szabo (Ed.), *The collected papers of Gerhard Gentzen* (pp. 68–131). North-Holland. https://doi.org/10.1016/S0049-237X(08)70822-X (Original work published 1934)

Gibbard, A. (1975). Contingent identity. *Journal of Philosophical Logic*, *4*(2), 187–221. http://www.jstor.org/stable/30226117

Gödel, K. (1932). Zum intuitionistischen Aussagenkalkül. *Anzeiger Akademie Der Wissenschaften Wien*, *69*, 65–66.

Godo, L., & Rodríguez, R. O. (2008). Logical approaches to fuzzy similarity-based reasoning: An overview. In G. Della Riccia, D. Dubois, R. Kruse, & H.-J. Lenz (Eds.), *Preferences and similarities* (pp. 75–128). Springer Vienna. https://doi.org/10.1007/978-3-211-85432-7_4

Goguen, J. A. (1968–1969). The logic of inexact concepts. *Synthese*, *19*(3/4), 325–373. http://www.jstor.org/stable/20114646

Graff, D. (2000). Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics*, *28*(1), 45–81. https://doi.org/10.5840/philtopics20002816

Hall, E. W. (1933). Numerical and qualitative identity. *The Monist*, *43*(1), 88–104. http://www.jstor.org/stable/27901391

Hawley, K. (2004). *How things persist*. Clarendon Press. https://doi.org/10.1093/acprof:oso/9780199275434.001.0001

Hume, D. (2010). *A treatise of human nature*. The Project Gutenberg. https://www.gutenberg.org/ebooks/4705 (Original work published 1739–1740)

Hyde, D. (1997). From heaps and gaps to heaps of gluts. *Mind*, *106*(424), 641–660. http://www.jstor.org/stable/2254669

Hyde, D., & Raffman, D. (2018). Sorites paradox. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2018/entries/sorites-paradox/

Jacas, J., & Recasens, J. (1992). Eigenvectors and generators of fuzzy relations. *Proceedings from the IEEE International Conference on Fuzzy Systems*, 687–694. https://doi.org/10.1109/FUZZY.1992.258741

Jacas, J. (1990). Similarity relations: The calculation of minimal generating families. *Fuzzy Sets and Systems*, *35*(2), 151–162. https://doi.org/10.1016/0165-0114(90)90190-H

Juti, R. (2001). *Johdatus metafysiikkaan* [Introduction to metaphysics]. Gaudeamus.

Klawonn, F., & Castro, J. L. (1995). Similarity in fuzzy reasoning. *Mathware & Soft Computing*, *2*, 197–228.

Kripke, S. A. (1980). *Naming and necessity*. Harvard University Press. (Original work published 1972)

Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, *2*, 458–508.

Lewis, D. (1986). *On the plurality of worlds*. Blackwell.

Lewis, D. (1988). Vague identity: Evans misunderstood. *Analysis*, *48*(3), 128–130. https://doi.org/10.1093/analys/48.3.128

Lewis, D. (1993). Many, but almost one. In J. Bacon, K. Campbell, & L. Reinhardt (Eds.), *Ontology, causality and mind: Essays in honour of D. M. Armstrong* (pp. 23–38). Cambridge University Press.

Locke, J. (2004). *An essay concerning humane understanding* (Vol. 1). The Project Gutenberg. https://www.gutenberg.org/ebooks/10615 (Original work published 1690)

Łukasiewicz, J., & Tarski, A. (1930). Untersuchungen über den Aussagenkalkül. *Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie: Cl. iii*, *23*, 30–50.

MacFarlane, J. (2010). Fuzzy epistemicism. In R. Dietz & S. Moruzzi (Eds.), *Cuts and clouds: Vagueness, its nature, and its logic* (pp. 438–463). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199570386.003.0026

Machina, K. F. (1976). Truth, belief, and vagueness. *Journal of Philosophical Logic*, *5*(1), 47–78. http://www.jstor.org/stable/30226134

Menger, K. (1951). Probabilistic theories of relations. *Proceedings of the National Academy of Sciences of the United States of America*, *37*(3), 178–180. http://www.jstor.org/stable/88063

Metcalfe, G., Olivetti, N., & Gabbay, D. M. (2008). *Proof theory for fuzzy logics*. Springer. https://doi.org/10.1007/978-1-4020-9409-5

Mostert, P. S., & Shields, A. L. (1957). On the structure of semigroups on a compact manifold with boundary. *Annals of Mathematics*, *65*(1), 117–143. https://doi.org/10.2307/1969668

Negri, S., & von Plato, J. (2001). *Structural proof theory*. Cambridge University Press. https://doi.org/10.1017/CBO9780511527340

Negri, S., & von Plato, J. (2011). *Proof analysis: A contribution to Hilbert's last problem*. Cambridge University Press. https://doi.org/10.1017/CBO9781139003513

Noonan, H., & Curtis, B. (2018). Identity. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2018/entries/identity/

Paganini, E. (2011). Vague objects without ontically indeterminate identity. *Erkenntnis*, *74*, 351–362. https://doi.org/10.1007/s10670-010-9257-8

Parsons, T. (2000). *Indeterminate identity: Metaphysics and semantics*. Clarendon Press. https://doi.org/10.1093/acprof:oso/9780198250449.001.0001

Parsons, T., & Woodruff, P. (1995). Worldly indeterminacy of identity. *Proceedings of the Aristotelian Society*, *95*, 171–191. https://doi.org/10.1093/aristotelian/95.1.171

Pelletier, F. J. (1989). Another argument against vague objects. *The Journal of Philosophy*, *86*(9), 481–492. https://doi.org/10.2307/2026760

Plutarch. (1914). *Plutarch's lives* (B. Perrin, Trans.; Vol. 1). William Heinemann.

Poincaré, H. (1958). *The value of science* (G. B. Halsted, Trans.). Dover. (Original work published 1905)

Priest, G. (1991). Sorites and identity. *Logique et Analyse*, *135-136*, 293–296.

Priest, G. (1998). Fuzzy identity and local validity. *The Monist*, *81*(2), 331–342. http://www.jstor.org/stable/27903588

Priest, G. (2003). A site for Sorites. In J. Beall (Ed.), *Liars and heaps: New essays on paradox* (pp. 9–23). Clarendon Press.

Priest, G. (2006). *In contradiction: A study of the transconsistent* (2nd, expanded ed.). Clarendon Press. https://doi.org/10.1093/acprof:oso/9780199263301.001.0001

Priest, G. (2008). *An introduction to non-classical logic: From if to is* (2nd, expanded ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511801174

Priest, G. (2010). Non-transitive identity. In R. Dietz & S. Moruzzi (Eds.), *Cuts and clouds: Vagueness, its nature, and its logic* (pp. 406–416). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199570386.003.0024

Quine, W. V. (1969). *Ontological relativity and other essays*. Columbia University Press.

Rosen, G., & Smith, N. J. J. (2004). Worldly indeterminacy: A rough guide. *Australasian Journal of Philosophy*, *82*(1), 185–198. https://doi.org/10.1080/713659795

Ruspini, E. (1982). Recent developments in fuzzy clustering. In R. R. Yager (Ed.), *Fuzzy set and possibility theory: Recent developments* (pp. 133–147). Pergamon Press.

Russell, B. (1923). Vagueness. *Australasian Journal of Philosophy and Psychology*, *1*, 84–92.

Sainsbury, R. M. (1995). Why the world cannot be vague. *The Southern Journal of Philosophy*, *33*(S1), 63–81. https://doi.org/10.1111/j.2041-6962.1995.tb00763.x

Salmon, N. U. (1981). *Reference and essence*. Princeton University Press.

Sider, T. (2001). *Four-dimensionalism: An ontology of persistence and time*. Clarendon Press. https://doi.org/10.1093/019924443X.001.0001

Smith, N. J. J. (2005). A plea for things that are not quite all there: Or, is there a problem about vague composition and vague existence? *The Journal of Philosophy*, *102*(8), 381–421. http://www.jstor.org/stable/3655679

Smith, N. J. J. (2008a). *Vagueness and degrees of truth*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199233007.001.0001

Smith, N. J. J. (2008b). Why sense cannot be made of vague identity. *Noûs*, *42*(1), 1–16. https://doi.org/10.1111/j.1468-0068.2007.00671.x

Sorensen, R. (2018). Vagueness. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2018/entries/vagueness/

Tarski, A. (1986). What are logical notions? (J. Corcoran, Ed.). *History and Philosophy of Logic*, *7*(2), 143–154. https://doi.org/10.1080/01445348608837096

Trillas, E., & Valverde, L. (1984). An inquiry into indistinguishability operators. In H. J. Skala, S. Termini, & E. Trillas (Eds.), *Aspects of vagueness* (pp. 231–256). Reidel.

Tye, M. (1990). Vague objects. *Mind*, *99*(396), 535–557. https://doi.org/10.1093/mind/XCIX.396.535

Unger, P. (1980). The problem of the many. *Midwest Studies in Philosophy*, *5*(1), 411–468. https://doi.org/10.1111/j.1475-4975.1980.tb00416.x

Valverde, L. (1985). On the structure of F-indistinguishability operators. *Fuzzy Sets and Systems*, *17*(3), 313–328. https://doi.org/10.1016/0165-0114(85)90096-X

van Inwagen, P. (1990). *Material beings*. Cornell University Press. http://www.jstor.org/stable/10.7591/j.ctt1jktrm2

Wiggins, D. (1968). On being in the same place at the same time. *The Philosophical Review*, *77*(1), 90–95. https://doi.org/10.2307/2183184

Williamson, T. (1994). *Vagueness*. Routledge.

Williamson, T. (2002). Vagueness, identity, and Leibniz's law. In A. Bottani, M. Carrara, & P. Giaretta (Eds.), *Individuals, essence, and identity: Themes of analytic metaphysics* (pp. 273–303). Springer. https://doi.org/10.1007/978-94-017-1866-0

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X

Zadeh, L. A. (1971). Similarity relations and fuzzy orderings. *Information Sciences*, *3*(2), 177–200. https://doi.org/10.1016/S0020-0255(71)80005-1