Vague Concepts and Sorites Paradoxes

Vlad Vieru

Machine Intelligence Research Laboratory
Ottawa
CANADA
e-mail:

"Whatever we may want to say, we won't probably say exactly that"
Marvin Minski

"It is... easy to be certain. One has only to be sufficiently vague."
Charles S. Peirce

"Subjectivity is the truth. By virtue of the relationship subsisting between the external truth and the existing individual, the paradox came into being."
Soren Kierkegaard

Abstract: The purpose of this article is the study of certain aspects regarding the indetermination introduced by vague concepts and, in particular, the explanation of sorites paradoxes. In the first part of the article, I will show that any theory on vague concepts must consider not only the relations between concepts and objects, but also the attitude of the users regarding these relations. Consistent to this point of view, the containing vague concepts will not be regarded as abstract descriptions. We will suggest instead an approach focusing on the individual descriptions belonging to various users at different moments in time. Based on this formal model and the associated semantics, we will show that sorites paradoxes are only apparently logic paradoxes. In fact, any sorites judgement speculates on our psychological difficulties in classifying objects versus one property or another.

1. The Nature of Vague Concepts

1.1 Terminology Used

Definitions

The term "user" shall refer to any human subject, as user of natural language and mental capabilities.

We will refer to something as *"(In)dependent of users"* if its existence and/or characteristics are (in)dependent of the users' perceptions/representations/attitudes of it. By (in)dependence we will mean the necessary, not the contingent.

"Real" or "*objective*" shall refer to any fact or thing completely independent of users. *Exterior reality* or *world* shall refer to all real things and facts.

"*Subjective*" shall refer to any fact or thing dependent on users. Accordingly, we shall refer to *interior reality* or *world*.

"*Objective concept*" shall refer to any concept whose intension is completely independent of users, while exclusively determined by the exterior reality. *Objective property* shall refer to a property which a certain thing really possesses or not.

"*Subjective concept*" shall refer to any concept whose intension is (partially) dependent on users. Accordingly, we shall refer to *subjective properties* or *predicates*.

When referring to "*users working (subjectively) with concepts*", we mean the associations which the users make (subjectively) between concepts and objects.

"*Positive extension*" and "*Negative extension*" of a concept shall refer to the multitude of objects to which the concept applies, respectively the multitude of objects to which the concept does not apply. By "*extension*", we implicitly mean "positive extension".

Remarks

1) Throughout the article, any person is implicitly considered a user.

2) Throughout the article the word "*concept*" will be used with preference. In some places, we will refer to *concepts*, although we would rather use conceptual *terms*. However, as tried to be suggested in Section 1.7, the distinction between concepts and terms designating concepts is not essential under the circumstances.

Similar reasons justify the flexible use of *"property"* and *"predicate"*.

3) The objectivity or subjectivity of a concept does not depend on the nature of objects to which the concept applies.

4) Although concepts have been classified with respect to their intensions, the objective or subjective nature of a concept is reflected in its extension as well. The extension of an objective concept is completely determined by the exterior reality, while the extension of a subjective concept is determined, at least partially, by the users' perceptions and attitudes.

5) We prefer to discuss objective concepts, and not *real* ones, in order not to leave the impression that we are approaching realism. From the very beginning, we consider that at least *certain* concepts (e.g. numbers, geometrical figures, physical measure units) are objective. For example, the mathematical concept of *triangle* is objective: something is or is not a triangle, regardless of anyone's opinion. This does not mean, however, that *triangularity* is an universal, in the sense used by the realist theories.

6) When we refer to certain properties or concepts as subjective, we almost always mean, even if not explicitly, that they are *partially* subjective. In other words, the intension of subjective concepts depends on the users' attitudes only to a certain extent. For some concepts, (e.g. the concept of *live being*), the users' subjectivity plays a small role. However, as long as a concept implies a certain dose of inherent subjectivity, we shall refer to the respective concept as being subjective. Our intention is not to defend relativism in this article, just to show that, to a certain degree, it is inevitable.

7) We will frequently use the term *"object"*. The meaning of the term varies according to the context in which it appears. Thus, *"object"* can be either an object *stricto sensu*, or a thing or fact making the object of a discourse or an intentional (i.e. faith, conviction, outlook, etc.). In its most general sense, *object* shall be any concrete or abstract entity (case, instance, moment of time) to which one concept or another can apply.

1.2 The Issue Proper

Together with the liar's paradox or Zenon's famous paradoxes, the sorites paradoxes are among the most lasting logic "games" ever imagined. Originally from ancient Greece, this type of paradoxes is still an object of joy and debate for many modern logicians. The name comes from the Greek word *soros*, meaning "heap". Actually, the best known sorites paradox remains the paradox of the heap. Initially, it was meant as an exercise: "Can one grain of wheat be described as a heap? No. Can two grains of wheat be described as forming a heap? No. ...Still, we must admit that, at one point, several grains of wheat do make a heap. The question is: when do several grains of wheat become a heap?". The heap paradox, as we know it today, conceals the same problem in the form of a logic reasoning:

"A grain of wheat does not make a heap.

If a grain of wheat doesn't make a heap,

Then neither do two grains of wheat.

If two grains of wheat don't make a heap,

Then neither do 3 grains of wheat.

If 9999 grains of wheat do not make a heap,

Then neither do 10.000 grains of wheat.

Therefore, 10000 grains of wheat do not make a heap."

The argument can go on forever, so that no number of wheat is enough to make a heap. Although the logic argument seems flawless, its conclusion violates the most basic intuition.

Formally speaking, the heap paradox can be transcribed as a logic argument with an unconditional premise ("1 wheat grain does not make a heap"), a general conditional premise ("If i wheat grains do not make a heap, then neither do $i+1$ wheat grains") and a conclusion like "10000 wheat grains do not make a heap" (10000 can be replaced with no matter how large a number). The only inference rules used are *modus ponens* and *cut*.

The formal aspects of the paradox will be discussed in detail later on. Right now, we shall only notice that the same type of reasoning can be applied to a wide range of cases, and is not specific to the heap paradox only. For example, in the same manner it can be shown that one person having 1 billion hairs is bald, that a person 2.20m in height is a tall person, etc. Paradoxes of this type form the class of sorites. The logical reasoning underlying them bears the generic name of sorites argument. At the same time, the sentences lending themselves to sorites arguments - for example, "x is a heap", "x is tall", "x is bald", etc. - are sorites predicates or sentences.

In the following, we shall be referring less to "sorites sentences" and more to "vague *concepts*". Although the two syntagms are quite close in meaning, the term "vague concepts" is more general. Discussing vague concepts does not imply an explicit reference to sorites paradoxes. However, the problem of vague concepts indetermination and that of paradoxes overlap to a large extent. This results clearly from the following interpretation of vague concepts:

"A term, or a concept, is vague if there are particular cases (instances, things, etc.) which do not fall clearly inside or outside the range of applicability of the term. At what point exactly does an object cease being red? How much hair exactly must a man lose to be judged bald? In what manner and to what degree must a person hold theological beliefs and practice the dictates of some religion to be regarded as religious? Such questions admit of no precise answers and the concepts redness, baldness, and religious must be regarded as vague." (Norman Swartz - "Beyond Experience").

Norman Swartz suggests as a (pseudo) definition of vague concepts the so-called notion of*"borderline cases"*. Borderlines are those cases in which the applicability of a concept cannot be clearly accepted or clearly rejected. For example, a person 1.75m tall represents a borderline case versus the concept of "tall person". A concept's shadowy area consists of all its borderline cases. According to several studies, a concept is vague to the extent it possesses borderline cases, namely to the extent to which its shadowy area is full. Any vague concept's main feature is the impossibility of drawing a firm line between its positive and negative extension. Thus defined, vague concepts coincide with what we called sorites sentences. The essence of sorites paradoxes is precisely the impossibility of determining the limits within which sorites sentences apply. Therefore any investigation of the nature of vague concepts will have as implicit result a certain interpretation of sorites paradoxes.

1.3 Types of Possible Solutions

Faced with sorites paradoxes, we can take one of the following four positions:

1) Deny the possibility of applying logic to sorites sentences

2) Deny one or more premises

3) Deny the logic reasoning validity

4) Accept the paradox

One answer to the sorites paradoxes evolves from the fundamental idea that the indetermination is of a linguistic nature. More logicians have suggested that, given the ambiguity and inaccuracy inherent to natural language, it goes beyond the purpose of logic to deal with linguistic structures as sorites paradoxes. According to this point of view, sorites paradoxes would result from an illegitimate attempt to apply the reasoning of classical logic to vague concepts. In order to eliminate the linguistic indetermination, a number of philosophers and logicians - among them Frege and Quine - proposed the creation of a so-called *ideal language* which would have to eliminate from the natural language any form of ambiguity and vagueness, enabling any phrase to be subject of logic investigations. Currently, the ideal language doctrine was practically abandoned. If we want logic to be a powerful enough instrument, then its application to the usual language must be possible. Therefore it is generally considered that logic can be applied to natural language, but that the sorites sentence is wrong, for one reason or another.

The type (2) answers accept classical logic and semantics as instruments of study, but consider the mistake to result from one or more incorrect premises. Since it seems impossible to deny the unconditional premise, the error consists in accepting a general unconditional premise. One of the most remarkable approaches of this type is known as epistemology. According to it, there would be a multitude of wheat grains not making a heap, but which would become a heap once *a single wheat grain is added* to it. Epistemologists state that there are precise dividing lines between heap and not heap, tall and not tall, red and not red, but that it is *us* who do not have the knowledge of distinguishing them. Therefore indetermination would be a form of ignorance and, accordingly, sorites paradoxes would not be logical, but epistemological. The main fault of the epistemological hypothesis resides in its counter-intuitive nature.

Most attempts to explain sorites paradoxes probably sprang from a type (3) position, namely from the feeling that there is something wrong with the logic reasoning or with the associated semantics. Generally, it is considered that the indetermination of vague concepts represents a semantic phenomenon. That is the reason why logicians focused on the attempt to revise classical semantics and with it the classical bivalent logic. The result was a relatively large number of non-classical logics (e.g. intuitionism, semantics supervaluationist semantics, paraconsistent logics) - also called non-standard or deviate logics. They generally operate with the idea that the simple notions of truth and false - as inherited from Aristotle - are not enough to approach the vagueness aspect. The solution suggested is almost always a multivalued logic (usually with 3 values). During the '60s and '70s, Zadeh and Goguen developed fuzzy logic with an infinity of truth values within the continuous interval of [0 1]. The main principle underlying fuzzy logic is the following: there is no absolute truth or falsehood, but degrees of truth. For example, just as a man can be bald to a higher or lower degree, so can the sentences to which the word "bald" applies have a higher or lower degree of truth. The adepts of fuzzy logic hold that the transition from a property to its opposite is generally continuous and gradual, hence it makes no sense to speak of an exact threshold between the extensions of the property (i.e. the positive and negative extensions). In particular, they assert that the transition from a non-heap to a heap takes place by a series of imperceptible qualitative changes. On a formal level, the paradox of the heap is explained by the fact that the word "heap" cannot be applied to a collection of i wheat grains in the same degree it can be applied to one of $i+1$ grains. Despite the claims to the solution to the problem of imprecision and the sorites paradoxes, as well as the fantastic success it enjoys, fuzzy logic begs a series of questions. For instance, it is not obvious that all statements that can be expressed in natural language may be compared as to their truth. What is more, fuzzy logic operates with an uncountable infinity of degrees of truth, while there is at most a countable infinity of possible sentences.

A last and desperate option is the unconditioned acceptance of the paradox. We can end up by accepting the idea that neither 10000 nor any other number of grains of wheat make up a heap. Setting aside the extremely counter-intuitive character of such a conclusion, we are left with an additional problem. Any sorites paradox admits a positive and a negative version. For instance, the paradox of the heap can be reformulated, in its positive version, as follows: *Ten thousand wheat grains form a heap. If 10000 grains form a heap, so do 9999. Finally, we reach the conclusion that a single grain of wheat forms a heap as well.* There exits no special reason for accepting the positive version of a paradox over its negative one, or the other way around. Therefore it must be concluded that words such as "heap", "bald", "tall" apply either everywhere or nowhere at all. In any case, their use is either redundant or self contradictory. Such a vision of natural language would be, we must admit, at least bizzare.

Taking exception from all these attempts to reformulate language, a small group of philosophers consider that objects themselves are vague. We want to exclude this metaphysical supposition from the scratch. Throughout the paper we shall use the fundamental premise that outside reality is precise. We do not believe there is a simple proof, or, indeed, any proof of this. Moreover, we do not know if the meaning of the statement *outside reality is precise* can be made perfectly clear. In any case, the premise of an outside reality does not play a crucial role in this paper. This is not essential for *supporting* the theory proposed, but for *refuting* the hypothesis of ontological indetermination as an explanation for sorites paradoxes. Therefore we do not believe that all readers believing in an imprecise reality should leave off at this point. I invite them to read further, at least those who do not believe that our problems in dealing with vague concepts are caused by a chaotic and vague reality.

As for the other approaches: we do not believe that the solution to the problem of indetermination or to the sorites paradox can be found in any of the previously mentioned ones. In our view, the "key" can be found only through an investigation that covers linguistic, semantic, logical, psychological and, not least, pragmatic aspects of the phenomenon.

1.4 Are Vague Concepts Really Vague?

In the previous Section we have explicitly underlined that objects are not intrinsically vague. What can be said, however, of vague concepts? *Are vague concepts really vague?*

Although rhetorically formulated, this question raises a very important point: is imprecision an essential feature of certain objects, or is it due to the way we *use* these concepts? In other words is a concept vague because it does not admit a precise application in certain cases, or because we do not know precisely how to apply it? Do borderline cases really exist, or are they just cases in which our imperfect perception of reality prevents us from discerning the applicability of concepts? Is classical logic really being tested, or is it just our ability to apply it to borderline cases? Is the extension of vague concepts really imprecise, or is it that we just do not know how to delimit it?

This series of questions could continue. But all these questions have, in the end, a common denominator. They question, on the one hand, the relation between concepts and objects, and on the other the role played by the user in this "equation". The following Sections will specifically address these problems. The declared goal is that, leaving aside the discussion of the epistemological hypothesis, the side effects of the investigation should reflect themselves in a plausible explanation of the phenomenon of imprecision.

1.5 Objective Concepts *Versus* Subjective Concepts

The *borderline* conception illustrated by the quotation from Norman Swartz does not point out the essence of imprecision, but only its symptoms. What does it really mean to say *"particular cases which do not fall clearly inside or outside the range of applicability of the term"?* This apparently unproblematic phrasing allows two distinct (though not disjoint) interpretations:

I_1: Users are unable to decide whether a certain vague concept can be applied in some special cases.

I_2: The possibility of applying a vague concept is inherently undecidable in some special cases.

Interpretation I1 only expresses an unquestionable fact. Besides being intuitively obvious, it is empirically confirmed by the following remarks referring to any vague concept:

1) at any time, for any limit case, some users consider the concept applicable, and others do not;

2) users change their opinion in time when dealing with limit cases.

These remarks prove, for now, only that *we* do not know how to establish the applicability of a concept to borderline cases. We can say that the way in which *we* apply vague words to objects is partially subjective.

But, leaving aside our incapacity, are the intensions and extensions of the concepts clearly delimited? Can we speak of a *real* applicability of concepts to things? Does it make sense to say that, in reality, any thing is either tall or not tall, either red or not red, etc. irrespective of our attitudes? The answer depends on whether we accept or reject the second interpretation, which states that imprecision resides in the intrinsic nature of the relation between concepts and objects, independently of the users.

The two interpretations are not independent, but the difference between them is essential. I1 could be considered a weak version of the interpretation of vague concepts, while I2 could be a strong one. The relation between them is $I_2 \rightarrow I_1$. The reciprocal implication is, however, invalid.

Indeed, if the applicability of vague concepts is, at least in some cases, meaningless, then users can deal with them only subjectively. But from our inability of finding an objective way of settling the applicability of a concept it does not follow that this applicability is not, in fact, uniquely determined by outside reality.

There is an analogy with the terms *decidable* and *undecidable*. When we wish to prove or refute a statement inside a formal system, it sometimes happens that we do not know how to do it (e.g. trying to prove Goldbach's conjecture). This does not mean that the statement is inherently undecidable in the respective formal system. On the other hand, if the system is incomplete, there are inherently undecidable systems. Any attempt to prove or refute such a statement inside the formal system is in principle doomed to failure.

Despite the analogy, there are remarkable differences. Problematic formal statements generally refer to the universality of a property - they state that a certain property is applied to every element of a certain infinite set. Our difficulty comes from the absence of a method to show that the property holds for *all* elements. Still, we can, in general, check the statement for a practically unlimited number of elements. When dealing with vague concepts, things stand differently. Not only do we not possess a recipe for their general use, but we do not even know how to test them objectively in borderline cases.

Returning to our two interpretations. I1 is practically a truism, I2 is debatable. We may choose to accept both, or just accept I1 and reject I2. Depending on our choice, we shall subscribe to one of two mutually incompatible points of view:

I_o: *Vague concepts are objective*

I_s: *Vague concepts are partially subjective*

The choice between these two alternatives, which we call the objectivist hypothesis and the subjectivist hypothesis respectively, holds, in our opinion, the key to understanding vague concepts and the sorites paradoxes.

One major difficulty is presented by the very classification of concepts in objective and subjective. The possibility of actually making this distinction and the criteria to use in making such a distinction are questionable. This is, in fact, the source of many open philosophical problems.

However, we should not draw the conclusion that any attempt to investigate the nature of vague concepts is doomed to failure. Quite the opposite: anyone interested in this question will answer it. But the answer, we believe, will not have universal validity. In last instance, any answer will represent the choice of one paradigm over another.

In the next Sections we shall attempt a personal choice for the subjectivist point of view.

1.6 A Pragmatic Argument in Favor of the Subjectivist Hypothesis

Before going into deeper arguments, we would like to point out the simplest argument in favor of the objectivist thesis, viz. "Occam's Razor": the more simple hypothesis is the likelier . In this case, given that users apply vague concepts subjectively, it is likelier that these are themselves subjective.

But in our opinion there are a lot more arguments in favor of the subjective thesis. In the following we shall present an essentially pragmatic argument, inspired by C. S. Peirce. Going against traditional metaphysics, Peirce says: *"It must be shown that almost every sentence of ontological metaphysics is either meaningless talk - in which a word is defined through another, and this one by others, without ever reaching a real concept - or is simply absurd."* Against such a sterile approach, Peirce suggested an "operational" theory of meaning, making practical verification into the criterion of truth. His point of view is that the entire meaning of an idea can be revealed by considering the practical implications of that idea.

"Our idea of anything is *our idea of its sensible effects; and if we fancy that we have any other we deceive ourselves, and mistake a mere sensation accompanying the thought for a part of the thought itself.* [...] *Consider what effects, which might conceivably have practical bearing, we conceive the object of our conception to have. Then our conception of those effects is the whole of our conception of the object."* (C.S.Peirce - "How To Make Our Ideas Clear")

In order to see how Peirce's principle can be applied to vague concepts, let us remember the weak version of their interpretation: *users do not know how to apply the term in some special cases.* This means that users *have no objective criteria* to determine the extension of vague concepts. To be more precise, this means that when considering the applicability of a vague concept to a borderline case, we are confronted with one of the following problems:

1) we cannot determine it *a priori*

2) we cannot determine it by empirical methods

3) we cannot establish it by an algorithm

We doubt anyone can seriously reject any of these problems.

The first states, for instance, that we cannot speak of *a priori* knowledge about the applicability of the concept "tall" to a man of height 1.75m. One should remark that we do not exclude the possibility of an interesting theory about tallness that starts by taking *any man of 1.75m is tall* as an *a priori* truth; we just do not think that such a theory could be accepted as a satisfying conceptual analysis of the word "tall".

The second remark is a generalization of the fact that no one has ever imagined - or will ever imagine - an empirical test to establish whether a 1.75m tall man is "tall". By empirical test we refer to observation and experiments performed on the object, not to the more or less accidental way that users report how objects seem to them relative to one concept or another. It is perfectly possible to conduct a survey among the users in order to determine their opinion on the applicability of a certain concept to a certain object at a particular time, but the results of this survey would show us not the relation between the object and the concept, but the attitudes of users to this matter. This paper attempts to show that investigating the attitude of the users is the *only* practical way of discussing the applicability of vague concepts, and that it is, therefore, the only one that is relevant. But for now, we are content with underlining that the extension of a vague concept cannot be determined solely by studying facts of the outside world.

Finally, in the last remark we take the algorithm to be describable using only precise concepts. It seems obvious that in this case there is no algorithm that can tell us whether, for instance, a 1.75m tall man is tall or not.

We have explained in detail what was meant by lack of objective criteria in order to point out two things:

1) the objectivist thesis stands no chance of a practical verification;

2) adopting the objectivist position can have no practical effect on the way *we* operate with vague concepts.

For these reasons, once we apply Peirce's principles the subjectivist hypothesis imposes itself. It is absurd to consider vague concepts objective as long as no user has objective criteria for their applicability.

1.7 Is *Red* An Objective or A Subjective Concept?

In the following we will try to anticipate and refute an objection that might be made against the previous argument.

When talking about the lack of objective criteria we took "tall" as an example. But, one might reply, there *are* objective criteria for some vague concepts, and therefore our generalization was faulty. Consider, for instance, the concept of *red.*

An objectivist could consider that applying a sorites paradox to *red* is wrong precisely because it presupposes *red* to be a partially subjective concept. The objectivist might support this by quoting from optical physics, which states that behind the sensation of *redness* lies a precise physical property: a wavelength between 620 and 760 micrometers. In other words, an object is either red or not red depending on the wavelength of the light it reflects. If this wavelength is 759 μm it is red, while if it is 761 μm it is not, no matter what the user thinks of it.

In our opinion, this line of argument is essentially flawed, because it supposes that the concept of red with which we operate in everyday life is, or should be, identical to the one used in optical physics. To show why this idea is false, we shall invoke a completely reasonable point of view:

"It is not essential that we try to get very clear what a concept is. That exercise may be left for books on the philosophy of language and of mind. Let me say only this: persons have a concept of - let us take as an example - redness, if they are able, for most part, to use correctly the word "redness" or some other

word, in some other language, which means pretty much what "redness" does in English." (Norman Swartz - "Beyond Experience").

Apparently, discussing concepts from the point of view of their use only troubles the waters. Objectivists will say that the use should be judged exclusively in reference to the state of facts in the outside world. On the other hand, subjectivists will hold that the standard of correctness in use is to be found not only in the outside world, but also in our perception of it. The weak point in Swartz's statement is that it assumes we know what the correct use of a word is, which is unfortunately not the case. It would seem that the problem has only been rephrased in terms of correct use of words.

In our opinion, the only way out is, again, judging the correctness of the use in a pragmatic fashion. In other words, give up the attempt to search, or impose, theoretical criteria independent of the users for the correctness of the use of words. The approach I suggest is summed up in the following pragmatic principle: *"Whatever works is correct."* Natural language works fine just the way it is. It fulfils its essential mission of facilitating the communication between its users. We can therefore assume that we use natural language correctly, at least most of the time.

Thus, the correctness of the use of words must be looked for in the capacity of the users to communicate something through them. This is independent of an agreement on what is being communicated. For instance, if one says a pie is tasty, everyone understands what is meant, irrespective of whether they agree that the pie is tasty or not. The concept of "tasty" is clear, what is questionable is its applications to particular cases.

Let us return to the color red. There are two different conceptions:

1) red designates a subjective visual experience;

2) red designates an objective physical property.

Let us take a user, say John, who says "I'm eating a red apple". His statement is perfectly clear. All users will understand that John's apple looks in a certain way. That is, the apple causes a specific visual sensation. This is the *only* plausible meaning of John's statement. It would be absurd to interpret John's message as "the apple I'm just eating reflects light with a wavelength between 620 and 760 μm". In fact, neither John nor anyone else need know anything about wavelengths at all.

We want to make ourselves perfectly clear. Even if everybody knew optical physics, no one had ever applied, or would ever apply, physical measurements in order to use "red" correctly. The property of wavelengths is simply irrelevant in understanding and using the phrase "red apple". *Red* is essentially a subjective sensation, not an objective property.

Correspondingly, the use of the word red in phrases such as "Hubble's red shift" supposes the understanding of it as an interval of wavelengths. In such a context, the interpretation of red as a visual sensation completely loses its relevance. Once we have decided to represent by "red" a set of wavelengths, it no longer matters whether we have or not a visual representation of red. The fact that, as it happens, we do have such a representation, affords a very intuitive interpretation. But the only meaning red has in such a specialized situation is that of a qualifier for a set of wavelengths.

It follows that in certain contexts we understand red to mean a visual sensation, and in others to be a set of wavelengths. There is, however, no context that requires us to understand red as both. The key to understanding red is to be found, always, in one interpretation or the other. We can safely say that, in this sense, the two meanings of red are *logically independent.*

As such, it is useless and counter-productive to refer to red as a unique concept. We can talk about two distinct concepts: *physical red* and *sensorial red* respectively. While physical red's extension can be clearly delimited from that of physical non-red, the same cannot be done for sensorial red. That is why using the wavelength definition of red does not solve the sorites paradox applied to colors.

1.8 The Failure of the Objectivist Thesis. Intermediary Conclusions

We have seen that physical measurements, no matter how precise, cannot quench our fascination with the impossibility of establishing an exact boundary for vague concepts. That is why the inherent indetermination in dealing with vague concepts, besides being logical or linguistic, is above all of psychological nature. Any attempt at solving the sorites paradoxes by examining only the outside world, or by reforming language, is doomed to failure. Objectivists can prove things about outside reality and about our theoretical possibility of representing it objectively, but they cannot prove that there is a unique way - lacking any imprecision - of *thinking*, that is, of representing reality in the *mind*. The simple fact that we are able to imagine vague concepts (such as tall, red, heap) which are inherently subjective and partially dependent on our mind is a proof of the objectivists' failure.

In conclusion, the objectivists have a flawed vision of vague concepts and as a consequence offer a false solution to the sorites paradoxes. In our opinion, any semantics of vague concepts must take into account not only outside reality, but also inner reality, i.e. the attitudes and conceptions of the users. In the second Part of this paper we attempt to show the implications of this fact at the level of logical formalism.

2. The Model of Subjective Descriptions

2.1 Introduction

Before studying the purely formal aspects of sentences about vague concepts, it is necessary to answer the following question: *what does it mean for a sentence to be true?* The Tarskian theory of correspondence-truth (*"The snow is white" if and only if the snow is white*) cannot be used here, since vague concepts are partially subjective and therefore it makes no sense to say, for instance, that "John is tall" is true if and only if John is *in reality* tall.

In the case of vague concepts, the criterion of truth defined by the relation between concepts and objects must be replaced by another: *the attitude of the users* regarding the relation between concepts and objects. To this end, we shall give a formal model built on the notion of *subjective description*.

2.2 Descriptions and Bets

(Pseudo-)definitions

A subjective description is a description that reflects the attitude of the user, at a certain moment, towards the applicability of a predicate to a specific object. Let P be a predicate: we call *P-description* any description referring to the relation between an object x and a predicate P.

Say that user u gives a positive (negative) P-description of object x at time t if u considers at time t that x has (has not) property P. Alternatively, say that x elicits a positive (negative) P-description from u at time t.

Remarks

1) Suppose that users decide the applicability of concepts to objects on the basis of bivalent logic, i.e. a description can be either positive or negative. This is simply a working hypothesis adopted for the sake of clarity. The formal model and its associated semantics can be extended with little effort to any multivalued descriptions.

2) Users give descriptions practically all the time. When saying "I am eating a red apple" one associates the object one is holding with the word apple, then describes the object apple by using the word red and finally associates the action on the thus described object with the verbal phrase "am eating".

3) There is no one-to-one relation between descriptions and sentences. Not every sentence expresses a real description, and not every mental description is communicated.

4) In order to gain an intuition of descriptions, we can imagine that any P-description given by a user u of an object x is equivalent to a bet on whether concept P applies to object x. We denote the bet by ?[P(x)] and the positive P-description by True[P(x)] and the negative P-description by False[P(x)]. We assume that u chooses the version that best reflects his attitude; if x is a borderline case, u does not have a well defined attitude towards either

description, but he *has to make a choice*. I take the moment u calls out his choice to be the moment the description is made. For instance, let us take as P the predicate "tall". If u gives a positive P-description of John, then he is betting on the sentence "John is tall", denoted by True[P(John)].

5) The betting interpretation of P-descriptions allows us to consider simultaneous P-descriptions. For instance, u can take part simultaneously in two bets ?[P(x)]and ?[P(x')]. The results of these bets will naturally satisfy certain compatibility conditions (for instance if John is sensibly shorter than Robin, it would be absurd for u to bet simultaneously on the sentences "John is tall" and "Robin is not tall").

Notations

Consider an arbitrary tuple (P, u, x, t). I shall use the following notations:

P (u, x, t) if u gives at time t a positive P-description of object x

~P (u, x, t) if u gives at time t a negative P-description of object x

#P (u, x, t) if u gives at time t no P-description of object x

Remarks

1) Using the betting interpretation:

   P (u, x, t) means that u bets on True[P(x)]

   ~P (u, x, t) means that u bets on False[P(x)]

   #P (u, x, t) means that u takes no part in bet ?[P(x)]

2) Notations P (u, x, t) and ~P (u, x, t) refer to actual descriptions, while #P (u, x, t) shows that it is meaningless to ask ourselves what u feels about applying P to x at time t. Actual descriptions are subject to the laws of bivalent logic (as per our assumption), but in order to faithfully describe the attitudes of users towards relations between concepts and objects we need this kind of trivalent logic.

For instance, the law of non-contradiction remains the same

~ [P (u, x, t) & ~P (u, x, t)]

But the law of the excluded middle is changed to

P (u, x, t) v ~P (u, x, t) v #P (u, x, t)

Definition

Two P-descriptions are identical P (u, x, t) = P (u', x', t') iff:

(P (u, x, t) & P (u', x', t')) v

(~P (u, x, t) & ~P (u', x', t'))

Otherwise the two P-descriptions are opposed:

P (u, x, t) $\neq$ P (u', x', t')

Remarks

1) In particular, it is possible to have: u = u' and/or x = x' and/or t = t'.

2) The identity of P-descriptions is not concerned with the identity of users or moments of time. For instance the descriptions expressed by "Milk is white" and "Snow is white" are identical because milk and snow are described in the same way relative to the predicate "white".

Definition

Object x elicits a *stable positive (negative) P-description* from user u if u never gives a negative (positive) P-description of x.

We use the notation:

P (u, x) if x elicits a stable positive P-description from u

~P (u, x) if x elicits a stable negative P-description from u

#P (u, x) if x does not elicit a stable P-description from u.

Remarks

1) Using the betting interpretation:

   P (u, x) means that u always chooses True[P(x)]

   ~P (u, x) means that u always chooses False[P(x)]

   #P (u, x) means that u does not always make the same choice.

2) Formally, the conditions for positive stability, negative stability and instability can be written as follows:

$$P(u, x) \leftrightarrow [(\forall t)\ (\#P(u, x, t)\ v\ P(u, x, t))]$$

$$\sim P(u, x) \leftrightarrow [(\forall t)\ (\#P(u, x, t)\ v\ \sim P(u, x, t))]$$

$$\sim P(u, x) \leftrightarrow [(\exists t)(\exists t')\ ((t \neq t')\ \&\ P(u, x, t)\ \&\ \sim P(u, x, t'))]$$

3) In general, the attitude of a user towards the relation between a predicate and an object changes in time. That is why P-descriptions are valid only for moments in time.Instead, stable P-descriptions are invariant in time.

4) An object may elicit a stable P-description from one user, but not from another. On the other hand, there are objects that elicit stable P-descriptions from all users. The next definition refers to these objects and their associated P-descriptions.

Definition

Object x elicits a universal positive (negative) P-description if no user ever gives a negative (positive) P-description of x.

We use the notation:

P (x) if x elicits a universal positive P-description

~P (x) if x elicits a universal negative P-description

#P (x) if x does not elicit a universal P-description.

Remarks

1) In the betting interpretation:

   P (x) means that no one ever bets on False[P(x)]

   ~P (x) means that no one ever bets on True[P(x)]

   #P (x) means that there exists at least one user who does not always choose the same bet as the others.

2) Formally, we can write the following relations:

$$P(x) \leftrightarrow [(\forall u)\ P(u, x)]$$

$$\sim P(x) \leftrightarrow [(\forall u) \sim P(u, x)]$$

$$\#P(x) \leftrightarrow [(\exists u)(\exists u')(\exists t)(\exists t')\ (P(u, x, t)\ \&\ \sim P(u', x, t'))]$$

3) Universal P-descriptions are specially important: they represent the formal instrument needed to avoid the trap of an absolute relativism. They are statements that transcend the quality of being simple unlasting opinions. Therefore, we could return to Tarski's truth as a correspondence theory, and say that truth is in the correspondence of the statement to reality. But we will also add an empirical criterion: a statement agrees with reality when there is a consensus on this matter. Once this pragmatic interpretation of truth is established, P-descriptions are equivalent to true statements. The sentence "10000 grains of wheat make up a heap" is no longer a subjective description reflecting an individual opinion, but a universally valid statement. The philosophy is : since everyone considers that 10000 grains of wheat make up a heap, that we can simply consider that 10000 grains of wheat do, *in fact*, make up a heap.

2.3 Dissonance and Indiscernability

Definition

We call two objects x and x' indiscernable relative to predicate P if whenever a user gives two simultaneous P-descriptions for x and x' they are identical. Otherwise, we call x and x'

discernable relative to P.

Remarks

1) Formally, the indiscernability condition is written as follows:

   $\mathrm{Ind}(P, x, x') \leftrightarrow \{ (\forall u) (\forall y, y \neq x, y \neq x') (\forall t) [\#P(u, y, t) \rightarrow (P(u, x, t) = P(u, x', t))] \}$

2) Two objects can be discernable relative to a property, but indiscernable relative to another. For instance, two twin brothers might be indiscernable relative to "tall" but discernable relative to "generous". This also serves to show that indiscernability is not always sensorial in nature.

3) Indiscernability is a binary tolerance relation (i.e. reflexive, symmetrical and transitive).

4) Indiscernability appears to be a vague notion. Supposing we define a distance between objects, we cannot find the maximum distance that satisfies:

   $(\forall x)(\forall y) [(\mathrm{dist}(x, y) < d) \rightarrow \mathrm{Ind}(P, x, y))$

Indetermination is not semantical, but epistemological. The *limits* of indiscernability cannot be set *a priori* because, although it is expressed as a logical relation, indiscernability is essentially defined on empirical grounds.

The axiom of indiscernable objects

If x and x' are two indiscernable objects relative to P, they cannot elicit opposite stable P-descriptions from any one user.

Remarks

1) Formally, the axiom of indiscernable objects can be written:

   $\mathrm{Ind}(P, x, x') \rightarrow \{ (\forall u) \sim[(P(u, x) \;\&\; \sim P(u, x')) \vee (\sim P(u, x) \;\&\; P(u, x'))] \}$

2) Although very natural, this axiom cannot be proved (there is no axiom that says that two indiscernable objects have to be considered simultaneously at least once).

Definition

An *empirically continuous series relative to a predicate P* is any ordered collection of objects X having the following property: any two consecutive objects of X are indiscernable relative to P.

Remarks

1) Let $X = \{x_1, x_2, \ldots, x_n\}$ be an ordered collection. The fact that X is an empirically continuous series relative to P may be written formally:

   $\mathrm{ECS}(X, P) \leftrightarrow [(\forall k, 0<k<n) \mathrm{Ind}(P, x_k, x_{k+1})]$

Definition

A *slippery slope* relative to P is any collection $X = \{x_1, x_2, \ldots, x_n\}$ satisfying the following two conditions:

i) it represents an empirically continuous series relative to P

ii) x_1 elicits a universal negative (positive) P-description and x_n elicits a universal positive (negative) P-description.

Remarks

1) Let $X = \{x_1, x_2, \ldots, x_n\}$. The fact that X is a slippery slope can be formally written as:

   $\mathrm{SS}(X, P) \leftrightarrow \{ \mathrm{ECS}(X, P) \;\&\; [(\sim P(x_1) \;\&\; P(x_n)) \vee (P(x_1) \;\&\; \sim P(x_n))] \}$

2) Sorites paradoxes are built on the grounds of a relation between objects forming a

slippery slope (the sorites argument is, in fact, also called the slippery slope argument). For instance, the series of quantities made up of one, two, ... ten thousand grains of wheat is a slippery slope relative to the property of being a heap.

Definitions

A P-classification of a set X at time t is the operation thereby a user classifies the objects of X at time t by predicate P.

A P-classification results in two sets X_1 and X_2 such that:

1) $X_1 \cup X_2 = X$

2) $(\forall x_k \in X_1) \quad P(u, x_k, t)$

3) $(\forall x_l \in X_2) \ {\sim}P(u, x_l, t)$

Subsets X_1 and X_2 are called equivalence classes relative to P.

Remarks

1) In the betting interpretation, the operation of P-classification is equivalent to taking part simultaneously in all the bets ?P[x], with $x \in X$. The two equivalence classes are made up of all the objects xk on which the user decides to bet True[P(x_k)], respectively of all the objects x_l on which the user decides to bet False[P(x_l)].

2) Formally, we can describe a P-classification of X by u at time t as:

   $(\forall x \in X) \ {\sim}\#P(u, x, t)$.

Definitions

Consider a user u and two objects x and x'. Supposing that u gives simultaneous P-descriptions for the two objects, we call the P-descriptions *dissonant* if the two objects are indiscernable. Otherwise, we call them consonant.

A P-classification of a set X is consonant if all P-descriptions of objects in X are mutually consonant. Otherwise, the P-classification is dissonant.

We shall say that u's attitudes at time t are consonant (or that u is consistent at time t) if all the P-descriptions given by u at time t are consonant. Otherwise, we shall say that u's attitudes at time t are dissonant (or that u is inconsistent at time t).

We use the notation:

Cons (u, t) if u is consistent at time t

~Cons (u, t) if u is inconsistent at time t

Remarks

1) In the betting interpretation, dissonant descriptions correspond to cases when the user bets simultaneously on True[P(x)] and False[P(x')], or False[P(x)] and True[P(x')]. In this case, u's bets are practically mutually contradictory.

2) The consonance condition can be formally written :

   $Cons(u, t) \leftrightarrow [\ (\forall x)(\forall y)\ (\ Ind(x, y)\ \&\ {\sim}\#P(u, x, t)\ \&\ {\sim}\#P(u, y, t)\ ) \rightarrow (\ P(u, x, t) = P(u, y, t)\ )\ ]$

3) The consonance condition is practically the principle of identity of indiscernable objects applied to simultaneous descriptions. The request for simultaneity is important, for there is no contradiction in applying a vague concept in a different way to a borderline case on two different occasions (in fact, as already discussed, this is an essential feature of vague concepts). It is important to distinguish between consonant attitudes and stable attitudes.

2.4 The Standard Version of the Paradox of the Heap

We now have all the formal instruments necessary to return to the sorites paradoxes. We shall start by stating once more the paradox of the heap:

"10000 grains of wheat make up a heap.

If 10000 grains make up a heap,

then so do 9999.

If 2 grains make up a heap,

then so does one.

Therefore one grain of wheat makes up a heap."

Let us denote by x_k a collection of k grains of wheat and by P the property of making up a heap. Formally, the paradox can be written:

$P(x_{10000})$

$P(x_{10000}) \rightarrow P(x_{9999})$

$P(x_{9999}) \rightarrow P(x_{9998})$

$P(x_2) \rightarrow P(x_1)$

===============

$P(x_1)$

The sorites argument is extremely puzzling, because of the evidence of the premises and the simplicity of reasoning, on the one hand, and the enormity of the conclusion, on the other.

Despite appearances, the sorites argument is not correct. Its error consists in implicitly adopting an objectivist interpretation of the concept of heap. When it says "k grains form a heap" it implicitly assumes that this statement is true independently of individual opinions. Or, we have seen that the truth as correspondence theory is not valid for vague concepts such as "heap". Vague concepts are partially subjective. We cannot consider that "k grains make up a heap" is an absolutely valid fact. "Heap" is, after all, only a label that someone attaches to a collection of grains at some time.

On the other hand, the formal transcription of the argument allows its interpretation by way of the model of subjective descriptions. The ordered collection $X = \{x_1, x_2, \ldots, x_{9999}, x_{10000}\}$ forms a slippery slope. By P(xk), we mean that all the users bet on $True[P(x_k)]$). Reading the sorites paradox in the model of descriptions does not eliminate the paradox, it just restates it. It continues to start from a true premise ("all users always describe 10000 grains of wheat as a heap") and reaches a false conclusion ("all users always describe one grain of wheat as a heap"). The paradox is built exclusively through logical relations between universal P-descriptions (a fact not surprising if thinking that universal P-descriptions represent the semantic equivalent of the objective truth).

In order to discover the logical flaw of the argument it is sufficient to analyze its fundamental premise:

If x and y are indiscernable relative to P and if x elicits a positive universal P-description, then y also elicits a positive universal P-description.

Using the betting interpretation this can be restated as:

If x and y are indiscernable and if all users always bet on True[P(x)], then all users always bet on True[P(y)]

No matter how counter-intuitive it may seem, this hypothesis is false. The statement will be rigorously accounted for, both empirically and logically, when giving the theorem of stability thresholds. For now, we will just remark that the classical version of the paradox is flawed by the assumption that any statement is or is not in agreement with reality, irrespective of what the users may think of. We can agree on this only there where precise conventions exist regarding the correspondence between certain concepts and the outside world. No such precise conventions exist for vague concepts. There are of course implicit rules governing the use of vague concepts which guarantee the viability of the communication between the users, but they are *ad hoc*, and their interpretation, at least for borderline cases, is left to the subjective appreciation of the users.

2.5 The Subjectivist Version of the Paradox of the Heap

The previous explanation, as good as it might appear, lacks an altogether convincing substance. Our feeling is that the sorites argument keeps inside a more general reasoning paradigm than ever guessed before. This fact made us venture not too far and keep close to what we called the *classical version* of the sorites paradox, i.e. an argument solely accountable in terms of universal P-descriptions.

We shall now try to study how the sorites reasoning may be extrapolated from universal

P-descriptions (i.e. *objective descriptions*) to arbitrary P-descriptions (i.e. *subjective descriptions*). This new form of the sorites argument was called the subjectivist version.

Consider, as above, predicate P and the slippery slope relative to P $\{x_1, x_2, \ldots, x_{10000}\}$. Additionally, consider a user u and a moment of time t. Assume that at time t, u wishes to simultaneously describe a grain of wheat and a collection of 10000 grains of wheat. If u describes the 10000 grains as making up a heap, then the subjectivist version "proves" that u will also describe one grain of wheat as making up a heap! Here is how the argument goes:

At time t, u considers that 10000 grains of wheat make up a heap

If u considers at time t that 10000 grains make up a heap, and if u is consistent at time t, then u considers at time t that 9999 grains make up a heap

If u considers at time t that 2 grains make up a heap, and if u is consistent at time t, then u considers that one grain of wheat makes up a heap.

Therefore either u considers at time t that one grain makes up a heap, or u is inconsistent at time t.

Although the subjective character of vague concepts is taken into account, the new version of the sorites argument reaches a conclusion no less paradoxical: in order to be consistent, a user cannot describe simultaneously two quantities of grains but identically, otherwise his attitudes are dissonant. This is obviously absurd. Try as we may, we could cannot find no dissonance in our decisions of describing a grain of wheat as not being a heap and 10000 grains as being one.

Any sorites paradox in subjectivist version uses the following premise:

"If x and y are two indiscernable objects, and u gives at time t a positive P-description of x, then u is consistent only if the P-description that u gives of y is also positive."

The betting interpretation runs as follows:

"If x and y are indiscernable objects, and u bets at time t on True[P(x)], then u is consistent only if he bets at time t on True[P(y)] "

Although it may seem correct, the premise assumes implicilty that if u gives at time t a P-description of x, then u must give simultaneously a P-description of y.

Nevertheless this hypothesis is wrong. Indeed, the fact that the user has at time t an attitude towards an object x is independent of whether the user has an attitude toward any other specific object. In particular, if u gives at time t a P-description of object x_k, then u may or may not give, at the same time t, a P-description of object x_{k-1} (i.e. the fact that u takes the bet $?[P(x_k)]$ does not entail the fact that he also takes the bet $?[P(x_{k-1})]$). All which consistency requires is that if u contingently has two simultaneous attitudes these must conform to certain compatibility conditions.

Assume that we have a grain of wheat on our right and 10000 grains on our left, when saying "there is a heap to my left and not a heap to my right" we only describe two entities. We have no attitude towards the application of the concept "heap" to other quantities of grains. In particular, we have no attitude toward the quantities $x_2, x_3, \ldots, x_{9999}$. At time t, only the problem of describing quantities x_1 and x_{10000} exists for us.

These remarks find a perfect correspondence at a formal level. The subjectivist version of the paradox of the heap is based on the following reasoning:

$\text{Cons}(u, t)$

$(\forall k, 1<k<10.000)\ \text{Ind}(x_k, x_{k+1})$

$P(u, x_{10000}, t)$

$P(u, x_{10000}, t) \rightarrow P(u, x_{9999}, t)$

$P(u, x_2, t) \quad \rightarrow P(u, x_1, t)$

========================

$P(u, x_1, t)$

Formally, the consonance condition is written:

$$\text{Cons}(u, t) \rightarrow [\ (\text{Ind}(x, x') \ \&\ \sim\#P(u, x, t) \ \&\sim\#P(u, x', t)\ ) \rightarrow (\ P(u, x, t) = P(u, x', t)\ )$$

The problem is that in the given situation the conditions $\sim\#P (u, x_k, t)$, $1<k<10000$ are not satisfied, so that the conditionals $P (u, x_k, t) \rightarrow P (u, x_{k-1}, t)$ are invalid. The P-description of object x_{10000} at time t does not have to be identical to the P-description of object x_{9999} at time t for the simple reason that, at time t, the user gives no P-description of object x_{9999}. Therefore *modus ponens* cannot be applied and the entire formal edifice crashes.

When u classifies a grain as not being a heap and 10000 grains as being a heap, he gives a P-description of the collection that holds only objects x_1 and x_{10000}. Since x_1 and x_{10000} are discernable relative to the concept "heap", their classification in opposite equivalence classes is quite consonant and natural. That is why the subjectivist version of the paradox of the heap is just as fallacious as the classical one.

2.6 The Classification of Slippery Slopes

In the previous Section we have shown that giving opposite descriptions to objects x_1 and x_{10000} implies no dissonance. But what happens if we want to simultaneously classify *all* objects of the set $\{x_1, x_2, \ldots, x_{10000}\}$? The answer is found in the following results.

Lemma

Any P-classification of an empirically continuous series is either dissonant or results in the assignment of all the objects to the same equivalence class.

Dissonance Theorem

Any P-classification of a slippery slope is inherently dissonant.

Formally, the theorem can be stated as follows:

$$SS (X) \rightarrow \{ (\forall u) (\forall t) [((\forall x \text{ e.} X) \sim\#P (u, x, t)) \rightarrow \sim Cons (u, t)] \}$$

We shall prove the theorem by *reductio ad absurdum*.

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a slippery slope such that $\sim P (x_1) \ \& \ P (x_n)$.

Let u be a user that at time t gives a P-classification of X.

Supposing Cons (u, t) we apply *modus ponens* for the slippery slope condition and for the consistency condition.

$$(\forall k, 1<k<n) (Ind (x_k, x_{k+1}) \ \& \ \sim\#P (u, x_k, t) \ \& \ \sim\#P (u, x_{k+1}, t))$$

$$(\forall k, 1<k<n) (Ind (x_k, x_{k+1}) \ \& \ \sim\#P (u, x_k, t) \ \& \ \sim\#P (u, x_{k+1}, t)) \rightarrow (P (u, x_k, t) = P (u, x_{k+1}, t))$$

==================================
==================================

$$(\forall k, 1<k<n) (P (u, x_k, t) = P (u, x_{k+1}, t))$$

We can now correctly apply the sorites argument.

$$P (u, x_n, t)$$

$$(\forall k) (P (u, x_k, t) = P (u, x_{k+1}, t))$$

=========================

$$P (u, x_1, t)$$

The conclusion reached, $P (u, x_1, t)$, contradicts the fact that x_1 elicits a universal negative P-description. Therefore the hypothesis Cons (u, t) is false.

q.e.d.

An analogous proof can be given for the slippery slopes having $P (x_1) \ \& \ \sim P (x_n)$.

Remark

If u is inconsistent at time t then u has to give at time t two dissonant P-descriptions. Therefore:

$$(\exists p, 1<p<n) (\sim P (u, x_p, t) \ \& \ P (u, x_{p+1}, t))$$

We call the corresponding pair of objects (x_p, x_{p+1}) the (subjective) *separation threshold* between the equivalence classes that result following the P-classification of series X.

The dissonance theorem is in full agreement with the personal experience of any one of us. Not only is there no objective, universally valid threshold between tall and not tall, heap and not heap, etc., but even establishing a subjective separation threshold gives rise to an inevitable internal inconsistency. Assume we wish to classify the various quantities of grains of wheat into heaps and non-heaps. Where can we draw the line? Between 599 and 600? Why not between 598 and 599 or between 600 and 601? The decision to be made will be eventually arbitrary, not only as pertaining to outside reality but also to our innermost convictions.

We do not want that the psychological issues of the dissonance theorem give the impression of justifying the adoption of an extremal skeptical or relativistic position towards vague concepts. Indeed, any classification of a slippery slope is arbitrary, but none is *completely* arbitrary. Although subjective separation thresholds vary from one classification to another – in relation to different users and different moments of time – they never exceed certain limits. These limits are the subject of the next two theorems.

The stability thresholds theorem

Let $X = \{x_1, x_2, \ldots x_n\}$ be a slippery slope such that $\sim P(x_1)$ & $P(x_n)$. Let u be a user that gives, at several moments of time, P-classifications of objects in X.

Then there is a stable pair of indiscernable objects $x_p, x_{p+1} \in X$ such that x_p does not elicit a stable P-description from u, but x_{p+1} elicits a stable positive P-description from u. Say that (x_p, x_{p+1}) represents the *individual positive stable threshold* of u. A similar definition holds for the individual negative stable threshold of u.

Remark

Under the betting interpretation the theorem reads:

There is a pair of indiscernable objects $x_p, x_{p+1} \in X$ such that u does not always bet $True[P(x_p)]$, but does always bet $True[P(x_{p+1})]$ (and similarly for the symmetrical case).

Formally, the theorem can be written:

$$SS(X) \rightarrow [\, (\exists p, 1<p<n)\, (\, Ind(x_p, x_{p+1})\ \&\ \#P(u, x_p)\ \&\ P(u, x_{p+1})\,)\,]$$

$$SS(X) \rightarrow [\, (\exists q, 1<q<n)\, (\, Ind(x_q, x_{q+1})\ \&\ \sim P(u, x_q)\ \&\ \#P(u, x_{q+1})\,)\,]$$

Suppose, as a *reductio ad absurdum*, that

$$(\forall k, 1<k<n) \sim [\, \#P(u, x_k)\ \&\ P(u, x_{k+1})\,] \quad (1)$$

Relation (1) can be rewritten

$$(\forall k, 1<k<n)\ [\, P(u, x_{k+1}) \rightarrow (\, P(u, x_k) \vee \sim P(u, x_k)\,)\,] \quad (2)$$

Now we can apply the sorites argument:

$P(u, x_n)$

(slippery slope condition)

$(\forall k, 1<k<n)\ [\, P(u, x_{k+1}) \rightarrow P(u, x_k)\,]$

((2) + indiscernability axiom)

=============================

$P(u, x_1)$

This conclusion contradicts the assumption that x_1 elicits a universally negative P-description. Therefore relation (1) is false and the existence of the positive stability threshold is proven. A similar proof can be given for the symmetrical threshold.

In order to make things clearer, we shall present the following example. Consider once again the collection $X = \{x_1, x_2, \ldots, x_{10000}\}$, where x_k represents *k* grains of wheat, and the concept of "heap" denoted by P. Imagine an experiment in which u gives, at different moments of time, P-classifications of different quantities of grains. Then, the theorem of stability thresholds asserts that there exist two indiscernable quantities x_p and x_{p+1} such that x_{p+1} is always classified as a heap, while x_p is at least once classified as not being a heap.

Assume that there are 100 P-classifications throughout the experiment. It is obvious that there are quantities u will describe as being a heap *exactly* 100 times, and others that u will describe as being a heap for *almost* 100 times. No matter what the result of the experiment is, there will always be two indiscernable quantities such that one is described as being a heap for 100 times, and the other for less than 100 times (for instance, for 99 times).

Let us now assume that the experiment consists of 1000 P-classifications. The previous statements remain valid. The only thing that changes is the effective position of the threshold. The number of P-classifications is immaterial. The more classifications, the farther to the right goes the positive stability threshold (i.e. the shadowy area of the concept of heap tends to grow). But no matter how long the experiment there is a limit that will never be exceeded. In other words, there exist two quantities x_p and x_{p+1} such that, *even if the experiment were to consist of an infinite number of steps*, x_{p+1} would always be described as a heap, and x_p would at least once be described as not being a heap. Since no user ever makes an infinite number of classifications, we can call (x_p, x_{p+1}) the *absolute* stability threshold of u.

In spite of the formal proof and the empirical considerations, we doubt that there will be many readers who will accept the existence of precise stable thresholds.The individual stability thresholds theorem induces its own kind of cognitive dissonance. The first temptation of a user is to argue as follows:"Assume the theorem holds and my own stability threshold is defined by the pair (x_p, x_{p+1}).But, since x_p and x_{p+1} are indiscernable quantities, how can I hold different attitudes toward them? If I have even once described x_p as not being a heap, why would I not ever make the same judgement for x_{p+1}?"

As in so many other cases this line of reasoning is confusing because it is based on a false hypothesis. In this case, it is incorrect to believe that if the positive stability threshold exits, then it can also be determined. In fact, any attempt u makes to identify through introspection his stability thresholds is doomed to failure. Whenever u tries to check whether (x_p, x_{p+1}) represent his positive stability threshold, he gives an *ad hoc* P-classification of the two objects. But since x_p and x_{p+1} are indiscernable objects, u holds at that moment identical attitudes towards them. Thus, u has no rational method of deciding on whether (x_p, x_{p+1}) is his positive stability threshold. The fact that a specific pair (x_p, x_{p+1}) defines the positive stability threshold is, after all, a chance matter.

This last remark might seem to open up the way for a probabilistic interpretation of the stability of descriptions. For instance, it might be considered that an object x elicits a positive stable P-description from user u of the probability that u gives a negative P-description of x be zero. Absolute stability thresholds would represent precise lines of delimitation between events having zero probability and events having non-zero probabilities of occurring. But such a purely theoretical probabilistic interpretation is inadequate. It would be impossible, for instance, to explain what "the probability that u describes 1 billion grains of wheat as not being a heap is *rigorously* zero".

If x is an object that elicits a positive stable P-description, then a negative P-description of x is not an event having a *mathematically* zero probability of occurring, but an event which only *contingently* never occurs. The existence of a negative P-description is not a possibility that is contradicted by theory, but one that is never confirmed by practice. As an analogy, let us imagine a potentially infinite number of throws of a die and the event that the face marked 6 comes up for 1000 times in a row. The probability that this event occurs is $1/6^{1000}$, not zero. This event may be a perfectly good example of an event that is *theoretically possible* but which never happens. We can well imagine a world in which someone throws the die and the face makes 6 comes up for 1000 times in a row. But in our world, this will never happen. Similarly, it is *theoretically* possible for someone to describe a quantity of one billion grains of wheat as not being a heap, but this will never happen.

Let us now suppose that the above experiment is extended to an arbitrary number of users. The positive stability thresholds will obviously vary from one user to another (although we would expect the difference between them be small). However, it is clear that there are two indiscernable quantities x_i and x_{i+1} such that x_{i+1} is always considered as being a heap by all the users, while x_i is for once, if not more, classified, at least by one user, as not being a heap. Therefore, we can extend the individual stability thresholds theorem to a universal stability thresholds theorem.

The consensus thresholds theorem

Consider a slippery slope $X = \{x_1, x_2, \dots x_n\}$ such that $\sim P(x_1)$ & $P(x_n)$.

There exists a pair of indiscernable objects x_p, $x_{p+1} \in X$ such that x_p does not elicit a universal P-description, but x_{p+1} does elicit a universal positive P-description. We shall call (x_p, x_{p+1}) the threshold between the shadowy area and the positive extension of P (an analogous theorem and definition holds in the symmetrical case).

Remark

Under the betting interpretation the theorem reads:

There exists a pair of indiscernable objects $x_p, x_{p+1} \in X$ such that not all users bet True[P(x_p)], but all users do bet on True[P(x_{p+1})] (and similarly for the symmetrical case).

Formally, the theorem can be written:

$$SS(X) \rightarrow [(\exists p, 1<p<n)(\, Ind(x_p, x_{p+1}) \;\&\; \#P(x_p) \;\&\; P(x_{p+1}))]$$

$$SS(X) \rightarrow [(\exists q, 1<q<n)(\, Ind(x_q, x_{q+1}) \;\&\; \sim P(x_q) \;\&\; \#P(x_{q+1}))]$$

The proof is similar to that given for the stability thresholds theorem.

2.7 Final Considerations

In general, logicians consider that the indetermination of vague concepts is recursive. Not only is the distinction between the positive and the negative extensions of a vague predicate blurred, but it is also unclear where this blurring starts and where it ends. In this sense, one talks about higher order indeterminations. The majority of studies treat exclusively the problem of order one indetermination. Instead , one of the declared strong suits of fuzzy logic is its claim of being a one shot solution to the types of indetermination.

The author, for one, considers that it does not make any sense to speak of a primary and a secondary indetermination. First order indetermination is of semantic character, and is manifested in the inherent subjectivity of deciding the separation threshold. Second order indetermination refers to the boundaries of subjectivity in establishing the separation thresholds. In our opinion this indetermination is mainly epistemological. Although we do not know the delimiting lines between consensus and subjectivity, they surely exist. These lines change with every new subjective that shatters the absolute consensus that exits at one time. Still, we can talk about absolute limits in the sense of the above absolute thresholds.

As a result, we can say that any vague predicate P determines a threefold divison of the universe of objects: the positive extension (objects to which P applies by absolute consensus), the negative extension (objects to which P , by absolute consensus, does not apply), and the shadowy area of P (objects for which there is no absolute consensus). The three regions correspond to:

$\{ x \mid P(x) \}$ – positive extension

$\{ x \mid \#P(x) \}$ – shadowy area

$\{ x \mid \sim P(x) \}$ – negative extension

The problem is that this threefold universal classification requires a higher point of vantage from which to analyze all the bivalent classifications made by each individual, which is practically an impossibility. Fortunately, in practice we rarely need absolute consensus and therefore we can use approximations.

Since we have admitted the practical impossibility of establishing a precise universal classification of objects in relation to a vague property, it might seem that indetermination cannot be done with by using the assumption of exclusively bivalent descriptions leading to a threefold universal classification. But this needs not favor multi-valued logics. First, the logical and psychological requirements of identically describing indiscernable objects do not depend on the number of available alternatives. Second, it is much more difficult to estimate the *degree of applicability* of a concept than to judge whether the concept is applicable or not. It is our opinion that any attempt at classifying objects in relation to a predicate in a higher (possibly infinite) number of classes of equivalence will not simplify the issue, but only make it more complex (i.e. it will not reduce the cognitive dissonance but rather amplify it).

BIBLIOGRAPHY

Ph. P. Wiener (Ed.) **Values in A Universe of Chance. Selected Writings of Charles S. Peirce**, DOUBLEDAY ANCHOR BOOKS, 1958.

CLARK, A., **A Physicalist Theory of Qualia**, THE MONIST, Vol. 68, No. 4, October 1985, pp. 491-506.

GOGUEN, J. A., **The Logic of Inexact Concepts**, SYNTHESE, 19, 1968, pp. 325-273.

HYDE, D., **Sorites Paradox** , Stanford Encyclopedia of Philosophy, 1996, http://plato.stanford.edu/entries/sorites-paradox/

KYBURG, H. E. and SMOKLER, H. E., **Studies in Subjective Probability** , JOHN WILEY & SONS INC., 1964.

NADIN, M., **The Logic of Vagueness and the Category of Synechism**, THE MONIST, Vol. 63, No. 3, July 1980, pp. 351-363.

POUNDSTONE, W., **Labyrinths of Reason: Paradox, Puzzles and The Frailty of Knowledge** , PENGUIN BOOKS, 1968.

RUSSELL, B., **Mysticism and Logic**, DOUBLEDAY ANCHOR BOOKS, 1957.

RUSSELL, B., **Our Knowledge of the External World**, THE NEW AMERICAN LIBRARY,1956.

RUSSELL, B., **Vagueness,** THE AUSTRALASIAN JOURNAL OF PSYCHOLOGY AND PHILOSOPHY 1, June 1923, pp. 84-92.

SORENSEN, R., **Metaphysics of Words**, PHILOSOPHICAL STUDIES, 81/2-3 March 1996, pp. 193-214.

SORENSEN, R., **Vagueness**, Stanford Encyclopedia of Philosophy,1997, http://plato.stanford.edu/entries/vagueness/

SWARTZ, N., **Beyond Experience**: **Metaphysical Theories and Philosophical Constraints,** UNIVERSITY OF TORONTO PRESS, 1991.

WITTIG, A. F. and WILLIAMS, G., **Psychology: An Introduction**, MCGRAW-HILL, 1984.