

# Revista Română de Filosofie Analitică

Romanian Journal  
of Analytic Philosophy

Volumul XV, Nr. 2, 2021



EDITURA UNIVERSITĂȚII DIN BUCUREȘTI  
BUCHAREST UNIVERSITY PRESS

2024



Revistă semestrială publicată de Societatea Română de Filosofie Analitică,  
cu sprijinul Facultății de Filosofie, Universitatea din București  
Revue semestrielle publiée par le Société Roumaine de Philosophie Analytique,  
soutenue par la Faculté de Philosophie, L'Université de Bucarest  
Biannual journal published by the Romanian Society for Analytic Philosophy,  
supported by the Faculty of Philosophy, University of Bucharest

Revista Română de Filosofie Analitică  
este fosta Revistă de Filosofie Analitică (2007-2011)  
© 2007, Societatea Română de Filosofie Analitică  
[www.srfa.ro](http://www.srfa.ro)

Ediția on-line: <http://www.srfa.ro/rrfa/>

**DIRECTOR / DIRECTEUR / DIRECTOR**

**Mircea Dumitru**, Universitatea din București

**REDACTOR-ȘEF / ÉDITEUR-EN-CHEF / CHIEF-EDITOR**

**Constantin Stoenescu**, Universitatea din București

**REDACTORI / ÉDITEURS / EDITORS**

**Andrei Mărășoiu**, Universitatea din București

**Bianca Savu**, Universitatea din București

**Andreea Popescu**, Universitatea din București

**Sandra-Cătălina Brânzaru**, Universitatea din București

**Bogdan Dumitrescu**, Universitatea din București

**Andreea Oprea**, Universitatea din București

**Rareș Dascălu**, Universitatea din București

**EDITOR ON-LINE**

**Lavinia Marin**, Katholieke Universiteit Leuven

**CONSILIUL ȘTIINȚIFIC / COMITÉ SCIENTIFIQUE / SCIENTIFIC BOARD**

**Marin Bălan**, Universitatea din București

**Radu J. Bogdan**, Universitatea Tulane, New Orleans

**Romulus Brâncoveanu**, Universitatea din București

**Cristian Calude**, Universitatea din Auckland

**Mircea Flonta**, Universitatea din București

**Mircea Dumitru**, Universitatea din București

**Radu Dudău**, Universitatea din București

**Kit Fine**, Universitatea din New York

**Adrian-Paul Iliescu**, Universitatea din București

**Hans-Klaus Keul**, Universitatea din Ulm

**Ilie Pârvu**, Universitatea din București

**Gabriel Sandu**, Universitatea din Helsinki

**Emanuel Socaciu**, Universitatea din București

**Wolfgang Spohn**, Universitatea Contanz

**Constantin Stoenescu**, Universitatea din București

**Ion Vezeanu**, Universitatea Pierre-Mendès-France, Grenoble

**REDAȚIA / SECRÉTARIAT / OFFICE**

Universitatea din București, Facultatea de Filosofie

Splaiul Independenței nr. 204, București, 060024

E-mail: [redactia@srfa.ro](mailto:redactia@srfa.ro)

**ISSN (ediția electronică): 1843-9969**

**ISSN (ediția tipărită): 1844-2218**



EDITURA UNIVERSITĂȚII DIN BUCUREȘTI

BUCHAREST UNIVERSITY PRESS

Bd. Mihail Kogălniceanu, nr. 36-46,  
Cămin A (curtea Facultății de Drept),  
Corp A, Intrarea A, etaj 1-2, Sector 5  
050107, București – ROMÂNIA  
Tel. + (4) 0726 390 815  
E-mail: [editura.unibuc@gmail.com](mailto:editura.unibuc@gmail.com)  
[www.editura-unibuc.ro](http://www.editura-unibuc.ro)

LIBRĂRIA EUB-BUP  
(Facultatea de Sociologie și Asistență Socială)  
Bd. Schitu Măgureanu nr. 9, sector 2  
010181 București – ROMÂNIA  
Tel. +40 213053703

TIPOGRAFIA EUB-BUP  
(Complexul LEU)  
Bd. Iuliu Maniu nr. 1-3,  
061071 București – ROMÂNIA  
Tel.: +40 799210566

DTP / COPERTĂ  
EUB-BUP

# Revista Română de Filosofie Analitică

Vol. XV

Iulie-Decembrie 2021

Nr. 2

## CONTENTS

LIA ZAMFIR, Cognition and consciousness in the yoga-samkhya philosophy – intersections with current debates in the philosophy of mind .....	73
MARIAN CĂLBOREAN, Vagueness and Frege .....	83
MARIA-FLORIANA GAȚE, Does the meaning of lying pose a problem to Pinocchio's paradox? .....	113



## COGNITION AND CONSCIOUSNESS IN THE YOGA-SAMKHYA PHILOSOPHY – INTERSECTIONS WITH CURRENT DEBATES IN THE PHILOSOPHY OF MIND

LIA ZAMFIR

**Abstract:** This paper explores the intersection of Yoga-Samkhya philosophy with contemporary debates in the philosophy of mind. While mainstream philosophy of mind has primarily embraced physicalism, asserting that everything has an underlying physical basis, it still fails to account satisfactorily for why or how exactly consciousness, and in particular its phenomenal aspect, would arise from neural structures and mechanisms. The paper argues for the relevance of ancient Eastern philosophies, specifically Samkhya-Yoga, in addressing persisting dilemmas regarding the relationship between the body, mind, and consciousness.

The Samkhya philosophy, rooted in metaphysical dualism between material (*prakṛti*) and spirit (*puruṣa*), introduces a perspective different from Western Cartesian dualism. The mind, in Samkhya, is considered part of nature or matter, distinct from consciousness or spirit. The article explores how Samkhya's ideas on the mind-body relationship, the role of intellect (*buddhi*), and the concepts of *gunas* (qualities of matter) and *koshas* (sheaths or aspects of human being) could enrich modern discussions.

Samkhya proposes a form of panpsychism, suggesting that consciousness is inherent in everything, while differentiating between an immaterial, immutable and unchanging witnessing consciousness (*puruṣa*), free from the constraints of physicality, time and space, and unaffected by the continuous fluctuations of the material mind.

The article concludes by asserting that considering Samkhya-Yoga's perspectives could reframe current debates, providing a constructive alternative to both Cartesian dualism and prevalent physicalism in understanding consciousness. The ancient philosophical insights from Samkhya-Yoga might offer valuable contributions to the ongoing discourse in the philosophy of mind.

**Keywords:** Samkhya-Yoga, Philosophy of mind, alternative dualism, panpsychism, consciousness.

## 1. Introduction

Most of contemporary philosophy of mind has so far been deeply entrenched in physicalism, which conceives of the world as essentially physical – meaning that all things and phenomena existing in the world are physical or have underlying physical processes – and strives to accommodate minds and consciousness in this physicality, by drawing on recent advances in biology, computer science, cognitive psychology, and neurosciences. However, in spite of the incredible pace of progress in these disciplines in the last decades, which enabled the empirical identification of ever more accurate correlates between behaviors, mental processes and neural (physical) activity, the widely-held assumption that this psycho-neural parallelism entails a causal or identity relationship remains an assumption, as there is no conclusive evidence for such causality and no satisfactory explanation for why or how exactly consciousness, and in particular its phenomenal aspect, would arise from neural structures and mechanisms. Some philosophers posit the impossibility of ever being able to give consciousness an adequate physical account. And if consciousness or any part of it cannot be accounted for in physical terms, this means that physicalism, the default position of modern science, is no longer a viable stand (Kim, 2011).

This paper aims to argue for the relevance of century-old ideas from Eastern philosophies, in particular the Indian Samkhya-Yoga schools of thought, for several of the ongoing controversies in the philosophy of mind.

The quasi-general lack of any reference to Eastern currents of thought in western Philosophy of Mind textbooks is astounding, given the intricacy and complexity of the models of the mind and consciousness they proposed.

It is our belief that considering these ideas can enrich and perhaps help advance the discussions on some of the persistent dilemmas on the relationship between the body, the mind and consciousness. They could potentially offer an alternative to the Cartesian metaphysics which, in



spite of being considerably more recent than these doctrines, is considered to have laid the cornerstone of the discipline of philosophy of mind, by formulating the mind-body problem, which remains, to these days at its center. In one way or another, all current views on the relation between physicality and a mentality inclusive of consciousness position themselves in some sort of response to the substance dualism asserted by Descartes. By formulating a consciousness-mind problem instead of the mind-body one, the Samkhya-Yoga could offer a change of reference point that might yield fresh insights.

## **2. A brief overview of the history and main tenets of the Samkhya and Yoga philosophies and the points of intersection with current notions in the philosophy of mind**

Samkhya is the oldest of the six orthodox *darshanas* or doctrines of Hindu philosophy, which were codified during the medieval period of Brahmanic-Sanskritic scholasticism, and which are called orthodox (*astika*) because they accept the *Vedas* as their scriptural authority. Its earliest written text, *Isvara Krsna's*, is assumed to date back somewhere between the 3rd and 5th century BCE. Samkhya is the doctrine with the greatest influence on the classical yoga theories and practices, as exposed by Patanjali in the *Yoga Sutras*, one of the earliest, and the most well-known treaty on yoga, compiling and systematizing ancient techniques, previously only subject to oral transmission through master-disciple initiation. In spite of the fact that Yoga, unlike Samkhya, does take a more theist stance, by adding a divine entity to Samkhya's 25 elements of reality, the two systems are strongly related and, as a consequence, are often treated together.

Without involving any concept of God, the Samkhya philosophy still advances a metaphysical dualism between material (*prakrti*) and spirit (*purusa*), while at the same time asserting the fundamentality of the latter. According to Eliade (1954/2009, p. 8), "The meaning of the term

Samkhya seems to have been “discrimination”, the chief end of this philosophy being to dissociate the spirit (*purusa*) from the matter (*prakruti*).” Unlike the Cartesian dualism that posits an ontological difference between mind and body, in Samkhya and Yoga philosophy, the distinction line is drawn between nature/matter and spirit, with both body and mind being considered to be part of the first. In that sense, similar to what Chalmers classifies as type F monism (Chalmers, 2003), this view shares some theses with both physicalism (postulating both body and mind to be simply different “states” or “densities” of matter) and substance dualism (claiming the existence of a transcendent, immortal self of pure and universal consciousness which underlies and is the ultimate source of the physical reality itself).

Consciousness, equated in Samkhya with this transcendent self, is stripped of most of the components it denominates in Western terminology and is narrowed down to pure spirit, who cannot receive any attributes or be in any relations, and of whom all one can say is “that it *exists* and that it *cognizes* (meaning, of course, that metaphysical cognition resulting from contemplating its own way of being)” (Eliade, 1954/2009, p. 12). Spirit is completely autonomous and inert, an indifferent and passive observer. But, while being clearly distinct from nature or matter, it is at the same time “intrinsic” to everything in existence, in a manner that is similar to the one in which Chalmers describes Bertrand Russell’s arguments for the existence of intrinsic properties of the fundamental physics systems:

“Perhaps the intrinsic properties of the physical world are themselves phenomenal properties. Or perhaps the intrinsic properties of the physical world are not phenomenal properties, but nevertheless constitute phenomenal, properties: that is, perhaps they are protophenomenal properties. If so, then consciousness and physical reality are deeply intertwined. This view holds the promise of integrating phenomenal and physical properties very tightly in the natural world. Here, nature consists of entities with intrinsic (proto)phenomenal qualities standing in causal relations within a spacetime manifold. Physics as we know it emerges from

the relations between these entities, whereas consciousness as we know it emerges from their intrinsic nature. As a bonus, this view is perfectly compatible with the causal closure of the microphysical and indeed with existing physical laws" (Chalmers, 2003, p. 37)

Samkhya and Yoga do not account for how the pure spirit became entangled with primordial matter in the first place, invoking an epistemological gap similar, again, to that suggested in the current day discussions of consciousness: grasping the cause and origin of the existing paradoxical relationship between spirit and nature is held to be beyond the current capability of human understanding. That is because human cognition is mediated by *buddhi* or the intellect, one of the constituents of the mind, which is nothing but a more subtle form of matter. The intellect's cognition is only limited to phenomena of the same ontological substance, therefore it cannot know the spirit or the conscious Self.

According to Samkhya and Yoga, *prakrti*, the metaphysical substance supporting all material or phenomena is also the substance of the mind. This substance homogeneity between the mind and the physical body eliminates the dilemma of mental causation, as the question of how two entities belonging to different realms could have effects on one another becomes irrelevant when both entities are seen as being part of the same realm.

*Prakrti*, the primordial matter has three different inherent "modes", qualities, or ways of being that are found in all its manifestations in variable and unstable proportions. These 3 different qualities of the matter, called *gunas*, are: *sattva* (the quality of translucence, intelligence, harmony), *rajas* (the quality of dynamism, transformation, striving) and *tamas* (the quality of inertia, stagnation, opaqueness).

The manifest matter is seen to consist of a mix of five elements: ether, air, fire, water and earth, each with different levels of density or subtlety (with the earth element being the densest and ether being the most subtle). In this taxonomy of the natural elements, ether is the controller of the embodied or experienced mind, called *antah karana*, sometimes translated as the inner instrument. The air/wind element is

what controls the pranas or energies of the body, the fire element is said to rule over the activity of all the sense and action organs (including the *antah karana*), while the sense organs themselves (eyes, ears, nose, mouth and skin) pertain to the water element. Finally, the action organs (speech, legs, hands, excretory and reproductive organs) are ruled by the earth element.

The body-mind system is experienced as five different *koshas* or *sheaths*: *annamaya* (the body of dense physical matter), *manomaya* (the body of mind, *antah karana*), *pranamaya* (the body of vital energies linking the gross and the subtle), *vijnanamaya* (the body of higher understanding or realization) and *anandamaya* (the body of beatitude or bliss). Experience of the five *koshas* is gradual and uneven (Niranjanananda Saraswati, 1993/2009, p. 37-40).

The five-kosha model could be seen as a possibility to account for the *why* or *how* of the mind-body *supervenience*, which asserts, in its strong version, that every mental phenomenon has a physical substrate or supervenience base (Kim, 2013).

This *antah karana* or experienced mind, which could be equated with the western term of individual mind, has four functional constituents called *ahamkara* (ego/sense of I), *manas* (the lower mind or organ of cognition), *chitta* (memory) and *buddhi* (the intellect).

- *Mannas* is the component of the mind responsible for perception, for organizing raw sensations coming from the sense organs and the senses and for carrying out the lower cognitive functions of primary analysis and decision.
- *Buddhi* (the intellect), in contrast with *mannas*, is the faculty of reason performing higher level cognitive functions requiring intuition, insight and reflection (Schweizer, 1993).
- *Chitta* (memory storage) is the seeing aspect, the one observing and recording “all gross, subtle and causal manifestation, and experiences of consciousness and energy” (Niranjanananda Saraswati, 1993/2009, p. 37). *Chitta* is inert and lacks all intentionality.

- And finally, *ahamkara* (ego) is the sense of I, it is what gives a definite identity to the body-mind complex, the generator of subjectivity; that appropriates and thus personalizes the experiences of the *manas* and *buddhi*.

The *ahamkara* (ego) component of the mind can perhaps be equated with the generator of subjective experience, what Nagel called “what it is like to be” in a certain state (Nagel, 1979), or to Ned Block’s notion of P-consciousness (Block, 1997).

But what is important to highlight here is that, in the Samkhya-Yoga view, the whole realm of *prakṛti*, mind and hence first-person experiencing included, is “inherently unconscious and thereby incapable of producing consciousness as an effect” (Schweizer, 1993, p. 848). So all mind states and all processes happening in the mind are inherently unconscious, or happening in the dark. By their nature, they cannot be in any relation with the conscious spirit which is, in its unchanging and immutable essence above all experience. The only way for experience to come under the light of consciousness is for that light, which is of a different ontological character, to be reflected in the intellect. Because *buddhi* (the intellect) is the most *sattvic* (refined, translucent, subtle) component of the mind, it has the capacity to reflect the light of the Self/conscious spirit and shine it on whatever contents are ascending to the intellect. It is only in this way, through the mediation of the intellect that the light of consciousness is shined on the contents of experience.

“Thus conscious thoughts and perceptual experiences take place when *buddhi* receives representational forms, both perceptual and conceptual, from *manas*, the organ of cognition (...). So *buddhi* receives cognitive structures from *manas*, and conscious “light” from *purusha*, and in this manner, specific mental structures are capable of being illuminated by an external source, and thereby these structures are able to appear conscious. But consciousness itself is entirely independent of the particular thought structures it happens to illuminate.” (Schweizer, 1993, p. 848)

It is therefore only the representational forms of objects, and not the objects themselves that can accede to the light of consciousness, because the objects themselves have too big a share of *tamasic* (gross, opaque) qualities to allow the light of consciousness to permeate them. This is a view that agrees to some extent with representational theories of consciousness, such as the higher-order perception (HOP) and higher-order thought (HOT) theories which posit that, to be aware of something requires that we represent that thing in the mind first, and that failure to represent it makes it remain in the dark (or, in Samkhya terms, outside the light of consciousness). According to Samkhya and Yoga, it is the light of consciousness that allows for objects to be known in *buddhi* (which is the component of the mind performing the higher level cognitive functions like insight and intuition). Access to *buddhi*, which could well be equated with the Global Workspace, is also mediated by attention or awareness, which in turn is directed by the *ahamkara* (ego aspect of the mind).

The intertwining of Consciousness and Materiality posited by the Samkhya philosophy could be paralleled to a form of dual-aspect panpsychism, which Chalmers defines as the view that “everything is conscious” (Chalmers, 1996, p. 216). Like panpsychism, which, according to David Skrbina, lies at the intersection of ontology and a theory of mind (Skrbina, 2017), Samkhya proposes at the same time a theory on the nature of things – claiming that consciousness is a prerequisite of all manifestation and an all-permeating constituent of everything in existence, and a theory on the nature of the mind – which is seen as separate from consciousness and belonging wholly to the realm of materiality. By operating a substance distinction between an immaterial, immovable and unchangeable witnessing “consciousness” (*purusa*) that lacks all intention and agency, and “awareness” as a quality or activity of the material mind, Samkhya may actually present a more palatable form of panpsychism, as it seems less counterintuitive to accept the pervasiveness of such a neutral consciousness than that of “mind” or “mentality”.

“Whereas awareness (*antahkarana*) (namely intellect, egoity and mind) is active, intentional, engaged and at every moment a reflection of subtle materiality; consciousness (*purusa*) cannot think or act and is not ontologically involved or intentionally related in any sense to primordial materiality other than being passively present. (...) Samkhya philosophy thereby rejects idealism without giving up an ultimately transcendent “consciousness”. It also rejects conventional dualism by reducing “mentalist” talk to one or another transformation of material “awareness”; and it modifies reductive materialism by introducing a unique notion of “consciousness” that is nonintentional and has nothing to do with ordinary mental awareness.” (Larson & Battacharya, 2014, p. 77).

On the other hand, the *triguna* character of primordial materiality (*sattva, rajas, tamas*) could account for differences in levels of sentience or awareness between various objects existing in the natural world. Thus, while consciousness is intrinsic in everything that exists, not every object would necessarily have to be sentient or aware or have any mind-like qualities, simply by virtue of its proportion of the three *gunas*. This would reconcile panpsychism with emergentist views of the mind.

### 3. Conclusions

This paper constitutes an attempt to highlight some points of intersection between certain understandings attained by the ancient Samkhya-Yoga doctrines and current strivings in the investigation of mind and consciousness.

I believe that, by offering an alternative to both the Cartesian dualism and the physicalism in which most current standpoints on the issue of consciousness are entrenched, Samkhya-Yoga could allow certain controversies to be reframed in more constructive ways.

Parallels can be drawn between Samkhya-Yoga and the panpsychic worldview that an increasing number of thinkers see as the only viable alternative to the current materialist or physicalist paradigm, which fails to provide a satisfactory account for consciousness.

## REFERENCES

- Block, Ned. "On a confusion about a function of consciousness", in N. Block, O. Flanagan, & C Güzeldere (Eds.), *The nature of Consciousness: Philosophical debates* (pp.375-415). Cambridge, MA: MIT Press, 1997.
- Chalmers, David. *The Conscious Mind*. Oxford University Press, 1996.
- Chalmers, David. "Consciousness and its Place in Nature", in Stephen P. Stich and Ted A. Warfield (Eds.), *The Blackwell Guide to Philosophy of Mind*. Blackwell Publishing, 2003.
- Eliade, Mircea. *Yoga: Immortality and Freedom* (W. R. Trask, Trans. from French). Princeton University Press, 2009[1954].
- Kim, Jaegwon. *Philosophy of Mind*. Third Edition. Westview Press, 2011.
- Larson, G. J., & Bhattacharya, R. S. *The Encyclopedia of Indian Philosophies, Volume 4: Samkhya, A Dualist Tradition in Indian Philosophy* (Vol. 842). Princeton University Press, 2014.
- Nagel, Thomas. "What Is It Like to Be a Bat?" *Philosophical Review* 83(4), pp. 435–450, 1974.
- Niranjananda, Saraswati. *Yoga Darshan*. Yoga Publications Trust, 2009[1993].
- Schweizer, Paul. "Mind/Consciousness Dualism in Sankhya-Yoga Philosophy". *Philosophy and Phenomenological Research*, 53(4), pp. 845-859, 1993.
- Skrbina, David. *Panpsychism in the West*. MIT Press, 2017.



## VAGUENESS AND FREGE

MARIAN CĂLBOREAN

**Abstract:** A constant of Frege's writing is his rejection of indeterminate predicates as found in natural language. This paper follows Frege's remarks on vagueness from the early "Begriffsschrift" to his mature works, drawing brief parallels with the main contemporary theories of vagueness. I critically examine Frege's arguments for the inconsistency of natural language and argue that the inability to accommodate vagueness in his mature ontology is mainly due to heuristic rules of thumb which Frege took as essential, not to a deep problem in his fundamental apparatus.

**Keywords:** Frege, vagueness, indeterminateness, precision, theory of definition

### Introduction

This study<sup>1</sup> grew from two questions. First, where does indeterminateness stand in the context of Gottlob Frege's philosophy and how does he justify his constant rejection of natural language on account of it? Secondly, can Frege's constant doubts be assuaged by recent theories of vagueness? These two questions can only receive interlocking answers, as the justification Frege provides for his rejection might need to be compared with what we learned from the debate on vagueness started during the 1960s and 1970s, incidentally by some of Frege's rediscoverers, such as Michael Dummett.

---

<sup>1</sup> The history of this paper predates (Călborean 2020) which contains a very condensed variant of it in Chapter 11.

As Frege never gave a positive theory of vagueness, there is a danger of introducing too many distinctions he would not have recognized. To avoid anachronism, the method of the paper is to follow Frege's early work in roughly chronological order, up to the Frege of after 1891, where I switch to discussing Frege's stance thematically, in relation to his mature ontology and semantics. Frege's fragments relevant to vagueness are often intermingled with fragments bearing on other topics, reason for which I try to follow his remarks closely and compare them, from place to place, with the main strands of the post-1960 philosophy of vagueness.

Upon analysis, both Frege's fundamental apparatus and common vague predicates survive, his rejection of natural language being unmotivated.

### 1. "Begriffsschrift" and the Sorites

The aim of Gottlob Frege's work "Begriffsschrift" (1997a) is to provide a core<sup>2</sup> symbolic language for laws of thought, which language is also called '*begriffsschrift*' or '*conceptual script*'. In the preface of the work, Frege speaks of *begriffsschrift* as being a formula language adequate to express those proofs which can be given by logical means alone. He arguably lists two conditions: first, the language should be able to express a complete chain of inference, so that nothing from intuition can matter to proof and, secondly, it should conserve the utmost precision of inferences and relations (1997a, 48).

Frege speaks of '*begrifflicher Inhalt*' (conceptual content) as consisting of those kind of entities between which such proofs arise (1997a, 49–53) and of which *begriffsschrift* would therefore make use. Against the tradition of Aristotelian logic, the conceptual content of such phrases as "S defeated P" and "P was defeated by S" is held to be the same.

---

<sup>2</sup> 'Core' in the sense that it could later be applied to all sciences by way of special signs so as to become a "single formula language" partially realizing Leibniz's project of *calculus ratiocinator* (Frege 1997a, 50).

Frege also introduces a '*judgement*' as being the assertion of truth about a conceptual content and a '*function*' as the invariant part of a unitary sub-expression<sup>3</sup> replaceable by some other symbol in its places. After introducing letters and logical symbols into the language, the latter including quantifiers and truth-functional operators expressed as graphical (most being vertical) connections between them, Frege gives nine axioms. Frege's system is a second-order predicate calculus, including what we now call propositional logic and first-order predicate logic. In the final part of "*Begriffsschrift*", he puts the system to work, proving some theorems of mathematical induction.

While not all philosophers would agree that there are such things as purely logical proofs, conceptual contents common in various linguistic expressions or functions separable "*in thought*"<sup>4</sup>, Frege's distinctions seem to make possible, by the end of "*Begriffsschrift*", a rigorous<sup>5</sup> analysis of mathematical induction, containing proofs which are general and important. While Frege's formulation of them makes use of unrestricted second-order quantification, it is recognized that what is now known as '*classical logic*' springs from Frege's "*Begriffsschrift*" and its unprecedented success in formalizing this kind of proofs<sup>6</sup>.

The topic of vagueness appears in this final part of "*Begriffsschrift*". Frege defines consecutively the notions of a property being hereditary in a sequence<sup>7</sup> (2002, 55), then the notion of an object following another in a

<sup>3</sup> Frege speaks of 'simple or complex symbol' (1997a, 67).

<sup>4</sup> The distinction object-function is fundamental to Frege's project and modern logic (Heck and May 2013, 835).

<sup>5</sup> Rigor is one of the main motivations of Frege's project, comprising the two conditions already noted of nothing coming into a proof unnoticed and of conserving truth, i.e., the possible syntactic verification of correct derivation but also a theory of definition (Frege 1960, XXI).

<sup>6</sup> For both points see Jean van Heijenoort's introduction to "*Begriffsschrift*" (Frege 2002, 1).

<sup>7</sup> Sequences, or in Beaney's translation, "*f-series*" (Frege 1997a, 75) are sets which satisfy  $\forall x (Fx \supset \forall y (f(xy) \supset Fy))$ . I remark that there is an obvious parallel with the principle of tolerance for vagueness (Călborean 2020, 22).

sequence<sup>8</sup> and then he arrives at the base proposition of mathematical induction<sup>9</sup>. Frege expresses it in words and adds an aside:

“We can translate (81) thus:

*If x has a property F that is hereditary in the f-sequence, and if y follows x in the f-sequence, then y has the property F*<sup>10</sup>

For example, let F be the property of being a heap of beans; let f be the procedure of removing one bean from a heap of beans; so that, f(a,b) means the circumstance that b contains all beans of a heap a except one and does not contain anything else. Then by means of our proposition we would arrive at the result that a single bean, or even none at all, is a heap of beans if the property of being a heap of beans is hereditary in the f-sequence. This is not the case in general, however since there are certain z for which F(z) cannot become a judgement on account of the indeterminateness of the notion ‘heap’”

This is the Sorites paradox. We see that the property of being a heap of beans seems hereditary in the sequence of one-bean subtraction, i.e., that the property of being a heap is not lost by removing one bean. But that’s not “the case in general” as Frege puts it, because the property of ‘being a heap’ is in some way problematic.

Many philosophers start to discuss vagueness by assuming the existence of borderline cases, those where it is unclear whether the property applies or not<sup>11</sup>. There is a parallel with Frege’s certain “z” above: Frege says that since the notion ‘heap’ is indeterminate, there are certain z where “F(z)” cannot become a judgement. The notion of unjudgeable contents, i.e., conceptual contents which cannot be asserted, is once more discussed in “Begriffsschrift”, namely when Frege states at #3 that contents such as “house” belong to it (1997a, 53), the heap of beans above being the second such example. But the proposition “Eleven beans

<sup>8</sup> Also known in Quine’s terminology as “proper ancestral” (Frege 2002, 59).

<sup>9</sup> Frege writes in a footnote “Bernoulli’s induction rests upon this” (2002, 62). Michael Beaney calls it “the key point of mathematical induction” (Frege 1997a, 77).

<sup>10</sup> Here Frege inserts his footnote concerning Bernoulli (2002, 62).

<sup>11</sup> This formulation is very close to Rosanna Keefe and Peter Smith’s (1996a, 2).

are a heap of beans" is quite different from "house". It seems like there is an easy way of saying why the latter cannot become a judgement, namely, it is not predicative, that is, capable (if turned into a judgement) of becoming true or false. But this is precisely Frege's point: the grammatical form of a truth-carrying expression does not guarantee that the expression is also logically truth-asserting. As van Heijenoort puts it "With these few remarks, Frege puts vague predicates outside logic" (1986, 32).

Let us give a common form of the Sorites paradox, covering Dummett's Wang's paradox (Dummett 1996, 99) too:

*IB (Induction basis): An object corresponding to a number  $x$  under measurement  $m$  has property  $P$ .*

*IS (Inductive step)<sup>12</sup>: If an object corresponding to a number under measurement  $m$  has property  $P$ , so does an object corresponding to the next / previous natural number under measurement  $m$ .*

---

*C (Conclusion): Objects corresponding to any number under measurement  $m$  have property  $P$ .*

We can take 'group of beans' as measurement and 'being a heap' as  $P$ , so that we get:

*IB: A group of two hundred beans is a heap of beans.*

*IS: If  $a$  is a heap of beans, a group only one bean short of  $a$  will be a heap of beans.*

---

*C: A group of zero beans is a heap of beans.*

---

<sup>12</sup> Mathematical induction is not necessary for the paradox, *IS* can be replaced with a finite series of modus ponens or conjunctive syllogism (Williamson 1994, 24).

Epistemicists such as Timothy Williamson deny *IS*: there is a number in the measurement where the property *P* does not apply, and that number is next to one for which the property does apply (Williamson 1996, 279). Other philosophers, fuzzy theorists among them, deny that the repeated application of the induction step conserves truth (Machina 1996, 200). In comparison, Frege chooses to generally deny the general applicability of the inductive step: he denies that such a predicate can always even be asserted. This means going further than needed. Frege could have gone epistemicist *avant la lettre* and deny that the property is hereditary in the sequence of bean-subtraction, saying that there is such a number *y* smaller by one than *x* so that *x* beans is a heap of beans and *y* beans is not. He could have thought '*heap of beans*' parallel to a sharp-boundary property such as '*natural number in the second dozen*'. The fact that he does not do so raises the question of whether and how he allows some numbers of beans to go through and others not.

Timothy Williamson writes that what Frege has in mind here is that while the notion '*heap*' fails to refer, some of its predications may still be judgements, because those sentences would employ the problematic words as *idioms*, that is, shortcuts or revelatory images based on context or previous experience, that secure truth or falsity to the proposition. He writes: "it is not a cartographer's job to explain why travelers with bad maps or none at all sometimes reach their destinations" (1994, 44). This assumes that we need to read in the early Frege of "Begriffsschrift" his later distinction of sense and reference although it is precisely in Frege's eponymous article for that distinction that he repudiates some main points of "Begriffsschrift"<sup>13</sup>. That is, Frege's first work did not mention conceptual expressions referring at all (Heck 2012, 21–22). He did not deny conceptual content to those predications of certain "*z*", even though the later Frege would deny reference to vague concept-words and, under some interpretations, sense as well (1997b, 178)<sup>14</sup>.

Let us take a step back and ask whether the inability of sentences containing vague predicates to become judgements is, for Frege, solely a

---

<sup>13</sup> Namely that claiming identity is a relation between names (Frege 2002, 20).

<sup>14</sup> See below at section 4.

matter of them not becoming true or false. He does not affirm this. Thus, an alternative is to remember that the analysis of mathematical induction rests on quantifying over properties. And this assumes that there is a common logical form of predicates, so that a symbol can represent them. But indeterminate predicates could be interpreted as exceptions to Frege's theory of sequences: they seem to be hereditary in a sequence, yet they also indicate as absurd the predication of their corresponding *C* in the soritical series above. Of course, this means that there is a range of "z" where there will be trouble, this trouble zone still manifesting itself as lack of truth value for associated judgements. But the deeper problem is that the natural language term does not conform to Frege's expectations of logic. Frege hoped that by removing all particular content irrelevant to validity of proof (2002, 7), a consistent kernel of thought would be revealed, but vague predicates belie it, by embedding *prima facie* logical relations which can be turned into a contradiction by the laws of the system. Therefore, Frege's aside on the Sorites paradox is an illustration of what he expects of logic. Indeterminateness fails the minimal rigor necessary for a formula language based on distinguishing functions and arguments in natural language, without any supplementary semantic or ontological characterization<sup>15</sup>. And vague predicates are not adequately captured in a formula language expressing pure thought, because the pure thoughts they embed lead to contradictions.

Therefore, my first conclusion is that Frege takes indeterminateness as going against the assumptions of his project: that logical relations can be extracted from language without contradiction. Of course, this raises the question of how to get a grip on what logic can be, for Frege, in relation to natural language and what I called *prima facie* logical relations embedded therein.

Before turning to that issue, note that 'Heap' could be understood as per the epistemicists, in having a precise border in centimeters<sup>16</sup>. Or be understood statistically, as I argued elsewhere (Călborean 2020), as

---

<sup>15</sup> In "Begriffsschrift" Frege speaks of functions as "expressions", a point Philip Jourdain was to call a "trace of formalism" (Heck and May 2013, 831–32).

<sup>16</sup> On how one can understand Frege as epistemicist, see Stephen Puryear (2013, 123–27).

applying both truthfully and falsely to separate groups of the same number of beans. In both cases, Frege's project would stay the same, the single difference being that what he treats as embedded logical relationships of 'Heap' should be nuanced<sup>17</sup>.

## 2. The relation between logic and natural language

There are times when Frege takes linguistic form as determinative of logical distinctions. The grammatical articles make the best example. Frege insists that the definite article marks the difference between objects and concepts up to the point of hypostatizing enigmatic objects corresponding to expressions of the form 'the concept "man"'<sup>18</sup>. He will introduce a special function " $\xi$ " to play the role of definite article, that of turning a concept into an object when appropriate, by way of his Axiom VI of his "Basic Laws of Arithmetic"<sup>19</sup>. He will also see the indefinite article as determinative of concepts<sup>20</sup> and the German subjunctive mood as determinative of indirect reference (1997i, 162), among many other.

On the other hand, Frege's main achievement is taken to be the revealing of a single logical form underlying various forms of natural language and distinct from them. As already noted, he shows that the subject-predicate distinction does not belong to logic. He also only uses

---

<sup>17</sup> Contrast van Heijenoort: "Ordinary language is somehow too weak to stand the stress of bivalence and should not be asked to bear up against the requirements of logical rigidity." (1986, 41).

<sup>18</sup> Functions are unsaturated, therefore Frege doesn't mix concepts with objects, making concepts non-referable. Because he cannot accept a definite-article language expression not being a name, or a name not having a referent, he insists that there are such objects standing for expressions of the form "the concept 'x'" (1997b, 174-77). Frege will later review passingly a suggestion that these objects could be somehow identified with the extensions of concepts, but will make nothing of it (1997h, 187).

<sup>19</sup> "Here, then, we have a substitute for the definite article of language, which serves to form proper names out of concept-words" (Frege 2016, 19).

<sup>20</sup> "As soon as a word is used with the indefinite article or in the plural without any article, it is a concept-word" (Frege 1960, 64).



truth-functional operators, ignoring shading, that is, performative aspects of language. He famously denies that pairs like “Men are mortal” and “Cicero is mortal” are of the same logical form, i.e., he distinguishes subordination of concepts from falling of an object under a concept (1997f, 81). He also distinguishes conceptual marks, under which a concept is subordinated, and which are properties for the objects falling under the concept, from second-level properties, characterizing concepts: “The number of planets is 7” does not mean that 7 is a property of planets and a conceptual mark of ‘*number of planets*’, but a second-level property of ‘*number of planets*’ (1960, 64).

Frege also defends his appeal to linguistic distinctions in “On Concept and Object” thus:

“... my own way of [basing logical rules on linguistic distinctions] is something that nobody can avoid who lays down such rules at all, for we cannot understand one another without language, and so in the end we must always rely on other people’s understanding words, inflexions, and sentence-construction in essentially the same way as ourselves.” (1997h, 184)

Frege then affirms that he’s not trying to give a linguistic definition to logical concepts, but only hints, appealing for that purpose to “the general feeling for the German language” (1997h, 184).

Does this mean that once apprehended, the linguistic priors of logical distinctions can be discarded as eliminable from the system? There are commentators that see Frege’s mature semantics as applying only to perfect formal languages (Dummett 1996, 109), so for them the answer would be affirmative. But Frege’s insistence on some linguistic devices, especially articles, is simply too strong to conform to this interpretation. Frege’s constant point of equilibrium was that, for a successful logical system, the conceptual distinctions should lead to successful treatment of logical argument, i.e., results justify the distinctions made. This may be seen as akin to the Rawlsian reflective equilibrium in which logical principles and treatment of particular language contexts are balanced so that maximal explanatory output is achieved. In “Begriffsschrift”, Frege rejected the judgeability of vague contents, but he did not deny that *being*

*a heap of beans'* is indeed a property, since it met his only criterion available: being separable in thought. As Frege develops a semantic theory and a strict theory of definition, ordinary language will come increasingly into attack, and he will constantly reject vague predicates. But Frege will also constantly employ and exemplify his arguments with ordinary-language examples, the latter never being outside his philosophical project.

After "Begriffsschrift", Frege formulates the aim of defining the concept of number and the foundations of arithmetic logically. In his first rejection of the Kantian synthetic nature of arithmetic judgements, he affirms that the realm of arithmetic is the enumerable, and the enumerable comprises anything, including:

"... inner mental processes and events and even concepts, that stand neither in temporal nor in spatial but only in logical relations to one another. The only barrier to enumerability is to be found in the imperfection of concepts. Bald people for example cannot be enumerated as long as the concept of baldness is not defined so precisely that for any individual there can be no doubt whether he falls under it. Thus the area of the enumerable is as wide as that of conceptual thought." (1997f, 80)

Frege writes that vague predicates such as '*bald*' are not enumerable, thus being imperfect concepts. Enumerability here means at least that there should exist such a number as the number of all individuals falling under the concept. But Frege's argument seems misleading, as he accepts in #54 of his later *Foundations of Arithmetic* that there are concepts which cannot be counted, those known as non-sortal concepts:

"We can, for example, divide up something falling under the concept "red" into parts in a variety of ways, without the parts thereby ceasing to fall under the same concept "red". To a concept of this kind no finite number will belong." (1960, 66)

One way to preserve Frege's argument against '*bald*' is to take him speaking instead of the concept '*bald people*' having the appearance of a so-called sortal concept but being in fact uncountable. This is similar to how we have read Frege's soritical discussion in the "Begriffsschrift":

natural-language terms embed opposite logical intuitions. Without further elaboration, it is unclear why this false appearance cannot be circumscribed. For example, why could the proposition “*A* is a bald person or *A* is not a bald person” not be true, as the supervaluationists hold, without committing oneself to the truth of any of the disjuncts (Fine 1975)? Completely excluding vague concepts from the realm of conceptual thought, as the quote above does, seems unmotivated. That being said, this kind of formulation becomes common in Frege’s later work.

### 3. Frege’s ontology and semantics

Since the characterizations of ‘*function*’, ‘*argument*’, ‘*concept*’ and ‘*predicate*’ are missing or incomplete in “*Begriffsschrift*”, Frege’s later works clarify them, taking functions as primary. The function will be defined, on the model of mathematical functions, as a mapping<sup>21</sup> of objects (first-level functions) or functions (for second-level functions) as arguments to objects as values of the function. Functions can be either one-place (monadic) or two-place (dyadic). Monadic functions that map their argument only to the truth-values (the True and the False) are concepts. Dyadic functions that map their arguments to the truth values are relations. The function (or concept or relation) is never an object, it is *unsaturated*. That is why a predicate letter is always written with at least one letter in parentheses, so as to indicate the empty places of the function. By saturation, that is, the coming together of a concept and object as argument, i.e., predication, a proposition is obtained (Frege 1997c, 130–48; 1997h, 181–93).

Frege also introduces a special kind of object standing in one-to-one correspondence with functions: the value-ranges. The reason for their introduction is Frege’s Platonism, insisting that numbers are objects,

---

<sup>21</sup> Frege cannot be said to offer this as a definition for ‘*function*’. His only definition is negative: that which is not object (1997c, 140).

leading to them being defined in terms of objects<sup>22</sup>. The “Foundations of Arithmetic” had introduced ‘*extensions*’ with that role, assuming in a footnote “that it is known what the extension of a concept is” (1960, 79). This means approximately the set of all objects falling under the concept, but Frege generalizes the idea in his “Basic Laws of Arithmetic”. It is tempting to see value-ranges as sets of ordered pairs containing every object in the domain and the value of the function at that object, but Frege defines ordered pairs in terms of value-ranges (Heck 2012, 10), which are introduced, controversially, by contextual definition<sup>23</sup>.

Twelve years after the publication of “Begriffsschrift” and twelve years before the second volume of his “Basic Laws of Arithmetic”, Frege splits “Begriffsschrift”’s conceptual content into ‘*Sinn*’ (sense) and ‘*Bedeutung*’ (reference) in his article “Function and Concept” (1997c)<sup>24</sup>. The reference of an expression is defined as the object or concept represented through the sense of the expression and that can have identity (for objects) or identity-like (for concepts) relations with expressions differing from it only in sense. The second approach to defining the reference is to identify it with the scientific objects or concepts underlying the expression<sup>25</sup>. The sense, as said, is defined as the mode of presentation<sup>26</sup> of the reference. This table results:

---

<sup>22</sup> The so-called “Caesar problem”, in the sense that an identity of objects should say what it is for two numbers to be identical (Frege 1960, 79).

<sup>23</sup> The “Basic Laws of Arithmetic” introduce value-ranges at #3 then re-examine them in #10 and #29-#32 (Frege 2016, 7). See Heck for a critical analysis (2012, 129–34).

<sup>24</sup> He will clarify the distinction in further works (Frege 1997i; 1997b).

<sup>25</sup> Based on such Fregean quotes as “The *Bedeutung* is thus shown at every point to be the essential thing for science” (1997b, 178), and “A concept-word must have a sense too and if it is to have a use in science, a *Bedeutung*.” (1997b, 180).

<sup>26</sup> Or, for some commentators *the mode of determination* (expression found in a similar context in “Begriffsschrift”), i.e., the way by which the true meaning (reference) is to be reached (Beaney 1997, 23).

<i>Type of expression:</i>	<i>Reference:</i>	<i>Sense:</i>
Propositions (in direct speech)	Truth-value object: the True or the False	The thought (e.g.: what is common in different-language translations)
Names (definite descriptions and proper names)	The bearer (corresponding real object)	Hidden description (debated)
Concept-words (general expressions)	The concept (unsaturated function)	Sense of the concept-word (debated)

Frege's commitment to objectivity leads him to hypostatize the objects of the True and the False as real. It also requires expressions to find a *reference*. If they cannot do that, it means that they only have senses, they are *bedeutungslos*. Fictional names such as "Pegasus" and "Nausicäa" have senses, but no references. So do all propositions containing such names.

Therefore, Frege's ontology contains unsaturated functions and objects, the latter comprising physical objects, truth-values, numbers, and value-ranges. Arguably, for the later Frege, *thoughts* and *senses* more generally may be accepted, as he affirms their objectivity (1997j, 325–45).

Frege's semantics works towards two seemingly opposite directions. First, the context principle, postulated as the second fundamental principle of his "Foundations of Arithmetic"<sup>27</sup> holds that a term has meaning only inside a proposition. But Frege also argues that we can only learn language by deriving the composite meaning from the meaning of the parts<sup>28</sup> and he gives, in his posthumous "Notes for Ludwig Darmstaedter" a fragmentary statement of the so-called *building principle* for both sense and reference. That is, he arguably says that the sense of a complex expression is built from the senses of its parts and that the

<sup>27</sup> "Never to ask for the meaning of a word in isolation, but only in the context of a proposition" (Frege 1960, XXII).

<sup>28</sup> "The possibility of our understanding propositions which we have never heard before rests evidently on this, that we construct the sense of a proposition out of parts that correspond to the words" (Frege 1997e, 320).

reference of a complex expression is built from the references of its parts (Frege 1997g, 364–65). Without deciding the matter, we'll note that compositionality will be one reason for Frege's rejection of vague predicates.

How is compositionality supposed to work? Take the proposition "The Earth is round". On the side of sense, one would say that the immutable thought expressed by the sentence has contributions from the senses of its parts (Frege 1997g, 364) and, also, that its sense is that the conditions under which the proposition has the Truth as reference are fulfilled (Frege 2016, 50). Those are, in a truth-conditional reading, the scientific propositions which should be true for the Earth to be round. As for the side of reference, the predicate concept-words '*is round*' refer to a concept, namely a mapping from any object to the True or the False. The name "Earth" refers to one of those objects, namely the Earth. Hence the application of the reference of the predicate to the reference of its argument results beautifully in the reference of the entire proposition, namely the Truth. Obviously, 'Earth' and 'is round' are common natural language terms, yet we seem to have precise scientific understandings of both. Can compositionality work when applied to natural language predicates without such an understanding?

#### 4. Frege's main objections

Having stated the distinctions above, we can now discuss the mature Frege's objections against indeterminate predicates. The best-known such fragment is in "Basic Laws of Arithmetic", #56, we can call it 'the completeness fragment'. Under the heading "Principle of completeness", Frege writes:

"A definition of a concept (a possible predicate) must be complete; it has to determine unambiguously for every object whether it falls under the concept or not (whether the predicate can be applied to it truly). Thus, there must be no object for which, after the definition, it remains doubtful

whether it falls under the concept, even though it may not always be possible, for us humans, with our deficient knowledge, to decide the question. Figuratively, we can also express it like this: a concept must have sharp boundaries. If one pictures a concept with respect to its extension as a region in a plane, then this is, of course, merely an analogy and must be treated with care, though it can be of service here. A concept without sharp boundaries would correspond to a region that would not have a sharp borderline everywhere but would, in places, be completely blurred, merging with its surroundings. This would not really be a region at all; and, correspondingly, a concept without sharp definition is wrongly called a concept. Logic cannot recognize such concept-like constructions as concepts; it is impossible to formulate exact laws concerning them. The law of excluded middle is in fact just the requirement, in another form, that concepts have sharp boundaries. Any object  $\Delta$  either falls under the concept  $\Phi$  or it does not fall under it: *tertium non datur*. Would, for example, the proposition "Every square root of 9 is odd" have any graspable sense if *square root of 9* were a concept without sharp boundaries? Does the question, "Are we still Christians", indeed have a sense if it is not determined to whom the predicate *Christian* can be truly applied and from whom it must be withheld?" (2016, 70).

Care should be taken after the first read. Frege's main purpose is not to discuss natural language reasoning, as often thought, although that is certainly important (Fine 1975, 279). It is to press the importance of complete definition in mathematics. Thus, the completeness fragment is followed by a detailed critique of *piecemeal definitions* given by Frege's contemporary mathematicians. What Frege understands by "piecemeal" is the habit of introducing and modifying new terms as one likes. He writes it

"... consists in providing a definition for a special case – for example, for the positive whole numbers – and putting it to use and then, after various theorems, following it up with a second explanation for a different case – for example for the negative whole numbers and for Zero – at which point, all too often, the mistake is committed of once again making determinations for the case already dealt with" (2016, 70).

### a) Definitions may be implicit

Therefore, the completeness fragment is a condensation of Frege's position on conceptual definition. It is followed by examples of ambiguous or misleading definitions in mathematics, exemplified with ordinary-language predicates, as an introduction to Frege's discussion of definitions given by Cantor and other mathematicians. How is this to be applied to our current use of language?

For Frege, the reason for which piecemeal definition is unacceptable in mathematics seems to be that one can define and redefine anything. But natural language may resist unprincipled redefinitions, if one assumes there are such things as linguistic norms which stops any one speaker from stipulating '*tall*' to mean whatever they want. Thus, we may suppose that speakers have some, possibly implicit, definitions of common terms, for Frege's argument to be relevant. They may acquire them on learning the language, to the same effect as the explicit – even if piecemeal – definitions of mathematical concepts. Then Frege is justified in drawing a parallel and saying that, if '*Christian*' neither applies nor does not apply to – say – a member of the Church of Jesus Christ of Latter-day Saints because of complex theological debate, it would not fit his definition of a concept. That is, mapping any object to the True or the False.

Therefore, the term '*Christian*' had been defined implicitly, yet not correctly, so it did not turn into a concept. Taken to the extreme, this implies something like Peter's Unger nihilism (2017)<sup>29</sup> in the vagueness debate: natural-language terms simply have no meaning (especially reference) because they do not have a correct (and consistent) definition, at least until Frege's *begriffsschrift* gives them such a rigorous definition (Weiner 2010).

This extreme interpretation is hard to square with Frege himself relying on natural language and, suggesting, as at the end of the fragment, that such predicates as '*Christian*' may under some circumstances be already acceptable. Implicit definitions may work differently from explicit ones. Therefore, we need to see *why* Frege insists on each (first

---

<sup>29</sup> For a discussion on whether Frege can be read as nihilist, see also Puryear (2013, 136–37)



level) concept being defined for every object as argument, and this also for natural language predicates.

### b) Securing referents and compositionality

Since his distinction between sense and reference, Frege argues that science primarily needs to secure referents, to avoid blind alleys:

“It seems to be demanded by scientific rigour that we ensure than an expression never becomes *bedeutungslos*; we must see to it that we never perform calculations with empty signs in the belief that we are dealing with objects. People have in the past carried out invalid procedures with divergent infinite series. [...] What rules we lay down is a matter of comparative indifference.” (1997c, 141).

The plain read of this fragment is a purely heuristic rule, to the effect that time saving in scientific work is preferable. What Frege has in mind is that a definition can introduce (or recognize) mathematical objects into being, but only if it is unambiguous: As he remarks in “Foundations of Arithmetic”, there is no problem with the concept of Infinite, as long as is non-ambiguous:

“Any name or symbol that has been introduced in a logically unexceptionable manner can be used in our enquiries without hesitation, and here our Number  $\infty$ <sup>30</sup> is as sound as 2 or 3” (1960, 97).

And he accepts contradictory definitions as well (Frege 1960, 87).

In the case of conceptual expressions, their possible referents are concepts, that is, functions mapping objects to truth-values. And, in the completeness fragment above, Frege held that any concept-words lacking values at any object of the domain, i.e., even for  $\odot$  (The Sun), do not correspond to such a concept. That is, first, because Frege insists that all

---

<sup>30</sup> This is the way Frege writes aleph-null, the cardinality of natural numbers.

well-formed formulas of *begriffsschrift* should have exactly one referent, and that is only assured if for any  $\Delta$ <sup>31</sup> the logical operators and functions with which it forms more complex expressions define what kind of *referent* results for the combined expression. Compositionality is thus stronger on the side of reference. Such a semantic structure should exist that indicates how the meaning (reference) of complex expressions is built from simpler forms, parallel to the syntactic structure created by the application of rules of inference. And in this semantic structure, the only contribution of parts is to the truth of the complex expression (Williamson 1994, 38). Any contextualism is incompatible with what Frege sees as scientific, namely a single domain of all objects of which any thought is immutable.

A related argument is that any lack of reference can propagate itself through the system, affecting a large number of cases. Frege points out that for any  $x$  if " $x+1$ " is *bedeutungslos*, " $x+1 = 10$ " will have no solution, so it will not refer to either the True or the False. It will be *bedeutungslos* as well, illustrating how concepts and functions move together (Frege 1997c, 141). If we accept the sentence "My 49-year-old uncle is a bald person" in our language, without it being true or false, then the meaning of "All bald persons are over 50 years old" is lost, because of the logical relations embedded by language. We can say that " $A$  is a bald person or  $A$  is not a bald person" could not be true and it could not be false if '*bald*' was indeterminate of  $A$ . That's because neither disjunct would stand for a truth-value, and '*or*' is truth-functional. Indeterminateness is extended in all directions by the rules of *begriffsschrift*. Only by not accepting indeterminate propositions, the law of excluded middle seems to apply; since " $C \odot$ " is a well-formed propositional formula, then it must be either true or false. But this is not the case in contemporary supervaluationism: we can define truth as super-truth, namely truth in all worlds with full valuations: all logical truths will then be super-true, including the law of excluded middle, even when both disjuncts are indeterminate (Varzi

---

<sup>31</sup>  $\Delta$  is any possible object. Heck writes "The term  $\Delta$  is not supposed to be a name in *begriffsschrift* at all: It is a formal device [...] subject only to the condition that it should refer to some object in the domain" (2012, 58).

2007, 647; Fine 1975). Frege did not distinguish derivation (the law of excluded middle: either A or not A) from semantics (what is now called bivalence: any proposition be either true or false). Thus, even if truth-functionality extended indeterminateness, logical laws could be saved.

### c) The Indeterminate and higher-order vagueness

Frege's insistence that a concept should be either true or false at *all* objects can be objected to as damaging to science: there are scientific cases where it is reasonable to reserve judgement. Division by zero is undefined. Therefore, any mathematical sentence containing division by zero is to be undefined as well. This may lead to a three-valued logic in which the truth tables will have Indeterminate for any complex expression if any of component expressions had the Indeterminate. Remark that we cannot take the disjunction of Indeterminate with its negation as true. Also, where three-valuationists like Michael Tye say that the conjunction of False with Indeterminate results in the False, we would still have Indeterminate, in order to circumscribe the Fregean scientific project (Tye 1996, 282). But it is hard to see how Frege would accept such an object as the Indeterminate: it lacks the timelessness mark of Frege's Platonism expressed in the introduction of the True and the False. Moreover, Frege tests truth-values as references of propositions by their ability to be substituted<sup>32</sup> by one another, and, in plain speech 'A man of 170 cm is tall' does not seem replaceable with 'A man of 300 hairs is bald', in all contexts.

Yet, a three-valued logic or supervaluationism<sup>33</sup> may be acceptable formally to Frege, even if anachronistic. They would conserve all the truths of the begriffsschrift at classical truth-values, and simultaneously separate all propositions in either Fregean (or classical) or indeterminate.

It is an open issue whether any theory assigning a precise truth-value, or a precise lack of truth-value can do justice to indeterminateness. As Williamson puts it, "To fail to stipulate a value is

---

<sup>32</sup> For both see (Frege 1997i, 159).

<sup>33</sup> For a discussion of Frege's supervaluationist bend see (Weiner 2010, 48)

not to stipulate that there be no value" (1994, 41). Indeterminateness consists in, maybe among others, being unable to say where the border lies. This is the phenomenon currently discussed as higher-order vagueness (Williamson 1994, 2): it being indeterminate whether a case is indeterminate or not. As we shall see, Frege's metaphor of figures on a plane hints at the problem. But Frege's work contains no discussion of higher-order vagueness, and he usually speaks as if sharp boundaries are right around the corner.

In "On Sinn and Bedeutung" Frege proposes securing a reference to the '*divergent infinite series*' which was mentioned above, by stipulating it as 0. This strategy, if applied *ceteris paribus* to ordinary-language predicates, would eliminate higher-order vagueness up until a new object appears, when the definition should be re-worked so that it covers it. Although Frege often states that functions should have one value at any possible object, the necessary redefinition may proceed, so to speak, in bulk, or by 'fields', as he writes: "Every widening of the field to which the objects indicated by *a* and *b* belong obliges us to give a new definition of the plus sign" (1997c, 141)<sup>34</sup>. There is some tension between this proposal to extend a definition by cases and his criticism of piecemeal definition cited above that arguably does the same thing without so much rigor, which tensions reinforce my claim that Frege's criticism of piecemeal definitions is heuristic. Note that piecemeal definitions, while unprincipled and prone to error, need not go wrong in all cases. Great mathematicians gave piecemeal definitions, while avoiding contradiction (van Heijenoort 1986, 34). On the other hand, as results in mathematical logic after Frege's time showed, there is no guaranteed way to extend a theory while keeping all its properties, so Frege's case-by-case approach is heuristic as well. Take Presburger Arithmetic which is decidable and complete but does not contain multiplication. The addition of multiplication will get you Peano Arithmetic – undecidable and incomplete.

In conclusion, Frege's argument against indeterminate predicates both ignores higher-order vagueness and assumes that ordinary speakers should re-negotiate their usage – across the entire linguistic community –

---

<sup>34</sup> Contrast van Heijenoort (1986, 32), where he goes against Frege's explicit words.

each time a new situation or class of situations appear, which seems unrealistic<sup>35</sup>.

#### d) The region metaphor and incompleteness

Let us try to fix a precise sense for “sharp boundary of concept” mentioned above by Frege in his now-famous<sup>36</sup> metaphor of “a concept with respect to its extension as a region in a plane”.

A first interpretation may go like this. Supposedly conceptual expressions are similar to differently colored areas, so that their color difference creates the boundary. That kind of boundary does not correspond to any real object. Suppose the points constituting the plane are the objects in the domain and those objects amenable to our concept-words cluster together. That is, if our concept-word is ‘*tall man*’, non-persons will be to the bottom, and persons will be the points of the plane from the top left to the top right, ordered by height in centimeters. And our concept-words are supposed to pick up, i.e., to color all points corresponding to tall people. Under this extensional view, if the color fades slowly into the background as Frege has it, we could say that vague concepts would lead to vague extensions. And it would be impossible to define a set without knowing whether any object is in it or not. Yet, there is an easy way out. The contemporary fuzzy theorist works with the concept of fuzzy set. She replaces extensions with a fuzzy set where a membership function associates a number on the real interval (from 0 to 1 – a degree) with each object (Machina 1996, 180). As with the three-valued approach discussed before, it is technically feasible<sup>37</sup>, because all valuations involving classical truth-values stay the same, even though

---

<sup>35</sup> van Heijenoort writes “...not that such an enterprise cannot be carried out, but rather that neither mathematics nor ordinary language proceeds thus” (1986, 37). I don’t think it can be carried out, at least for mathematics.

<sup>36</sup> Boundaries are of course an ubiquitous metaphor in the debate on vagueness (Keefe and Smith 1996b)

<sup>37</sup> I do not affirm that the whole begriffsschrift could be conservatively extended, I mean only first-order predicate logic, excluding the second-order quantification.

the objection related to higher-order vagueness will apply with the same strength<sup>38</sup>. Since “exact laws concerning them” would still be possible, contrary to Frege, this would be to simply understand Frege’s stance as unimaginative of this further logical development and give up on it.

Another, more charitable explanation of the metaphor, rejects our ordering of persons according to their height, since nothing in fact requires that the area of the concept be compact. Frege does not indicate that indeterminateness is in any way regular or gradual<sup>39</sup>. Then, we should simply color those points for which the conceptual expression has the True as value. When we cannot do it, the blurring into the background would simply express our indecision, not a gradual decrease in height that could be correlated with a decreasing fuzzy value or probabilistic verity (Edgington 1996, 302). Therefore, Frege’s indeterminateness may not be gradual and, as we saw above, may not be equivalent to introducing a third truth-value, at least for those predicates displaying higher-order vagueness.

Williamson writes that Frege briefly compares blurred borders with dashed (interrupted) borders, without making much of it (1994, 279). Indeed, two types of indeterminateness should be distinguished against Frege<sup>40</sup>. The first is vagueness proper, characterized by higher-order vagueness, a problem Frege’s writings do not address. The other is incompleteness of definition. Namely, the definition of a conceptual expression is incomplete when we have some conceptual marks of the concept-words, but we have no ground to expect either that there are no other such marks or that there are. Metaphorically, the boundaries start strict, but we have no ground to foresee how or whether they continue. I

---

<sup>38</sup> A common philosophical objection to fuzzy logic is that it replaces vagueness with the maximum precision (Keefe and Smith 1996a, 46).

<sup>39</sup> Kit Fine introduced such rules under the name of “penumbral connections” or “truths on a penumbra”. Vague concepts are to be governed by some rules along the lines of “any blob redder than a pinkish blob is red if the latter is” (Fine 1975).

<sup>40</sup> “Frege’s requirement of completeness is intimately connected with that of sharpness. For him, in fact, the two requirements seem to fuse into one. Countless times in his writings, we find the words ‘complete’ and ‘sharp’ conjoined” (van Heijenoort 1986, 37).

think it is best to cite Fine's example, reminiscent, as he writes, of Carnap's meaning postulates:

- (a)  $n$  is nice<sub>1</sub> if  $n > 15$
- (b)  $n$  is not nice<sub>1</sub> if  $n < 13$  (Fine 1975, 266)

Frege gave himself one example of such incompleteness of definition: the Homeric  $\mu\omega\lambda\upsilon$  ("mōly"), a magic plant characterized by Michael Beaney as "having a black root and a milk-white flower" (Frege 1997b, 178). Frege says of it that it is *bedeutungslos*, although it is true that certain marks are supplied (*idem*). Let us remark that this seems to break the definition of conceptual marks given by Frege earlier, in that the proposition "All  $\mu\omega\lambda\upsilon$ s are black rooted plants" cannot be true as long as  $\mu\omega\lambda\upsilon$  is *bedeutungslos*. What we see is that Frege refuses to take the two conceptual marks he knows as the only ones. If he did so, ' $\mu\omega\lambda\upsilon$ ' would be a concept defined for all objects<sup>41</sup>, and thus sharply defined. Still, ' $\mu\omega\lambda\upsilon$ ' does not illustrate the same higher-order indetermination that '*bald*' does, so it would not be vague in the same sense, it is just incomplete.

A third interpretation of the region metaphor starts from the fact discussed above that definitions and redefinitions of concept-words work in bulk, based on the properties of objects. Not each one is to be assigned individually. As Williamson puts it: for Frege "to grasp as sense is to know where its boundary runs" (1996, 276). But the fragment above allows that "it may not always be possible, for us humans, with our deficient knowledge, to decide the question. "So the sharp boundary can be identified with the conceptual distinctions known, most probably by the scientific community as a whole or ideally, possibly on an extended timescale. Frege says in his "Foundations of Arithmetic" that if the manner of determining the pieces covered by a name changes, the objectivity of the determination will not. His example is that the objectivity of the North Sea will "not [be] affected by the fact that it is a matter of our arbitrary choice which part of all the water on the earth's

---

<sup>41</sup> At least assuming the concept-words expressing the two conceptual marks correspond themselves to real concepts.

surface we mark off and elect to call the ‘North Sea’” (Frege 1960, 34). This means again that as long as it is unambiguous, any definition will do, even if it is not presently used by the linguistic community. Then, what the metaphor shows is that the scientific community should know now or should be able to know in the future exactly what separates one concept from the rest, that potential criterion being serving itself as the sharp border of the metaphor, which is now blurred.

This third interpretation turns as well against Frege. He may not believe contemporary demographics a science, but there are contemporary surveys that count the numbers of Christians per country. Since Frege is committed to science ideally, there seems to be no a priori reason why such terms as ‘*Christian*’ cannot at one time receive a definition to serve as a sharp border, be it by self-report. Such definitions, even for our current use of ‘*Christian*’ may already be available, although not yet *discovered*. As stated, to bring rigor, Frege treated heuristic issues as constitutive. So, there is no wonder that the idea of *discovering* adequate definitions of natural language predicates did not arrive at him. This can be called Frege’s third heuristic: that criteria should be given, not waited for to be discovered.

#### **e) Do indeterminate expressions have a sense?**

Frege asks the final question of the completeness fragment as if it may be possible for the predicate “*Christian*” to have sharp boundaries. Were those missing, the proposition “*Are we still Christians*” would have no graspable sense. We can even strengthen Frege’s stance with some examples inspired from the contemporary debate on vagueness. Think of the persons of which most cannot say, and even most of them themselves cannot say whether they are *Christian*; they may identify themselves as “*cultural Christians*”, since they admire churches on the outside but never go in. Or, even though Frege himself wrote in his “*Foundations of Arithmetic*” that “*The number belonging to the concept ‘inhabitant of Germany at New Year 1883, Berlin time’ is the same for all eternity*” (1960, 60), we may suppose some German inhabitants in train of becoming



foreign residents at about 12am on the 1st of Jan 1883, German inhabitants having their last seconds of life around that time or stateless persons lost on the mountainous Swiss border<sup>42</sup>.

Let us now ask how would the presumed existence of some objects which are neither Christian nor not-Christian affect our ability to grasp the sense of the question “Are we still Christians”? That sense is to be a Fregean thought, and Frege tells us that even fictional thoughts can be grasped. He then may say that, since ‘Christian’ is not well-defined, it does not have a sense, so the whole thought does not have a sense.

But suppose two propositions:

*EP*: We are Elvish

*CP*: We are Christian

‘*Elvish*’ is fictional, thus no object falls under the concept. Suppose ‘*Christian*’ is identical with ‘*Elvish*’, except that for  $\odot$  and some other planets we fail to stipulate whether they fall or not under the concept. According to Frege *EP* has a sense (while being *bedeutungslos*) and so should *CP*, since the difference does not affect the thought. By this I mean that there is no connection between whether some objects are stipulated to fall under a concept or not and my grasping<sup>43</sup> of the words of *CP* as a mode of determination (i.e., towards whether *CP* is true or false). In his “Introduction to Logic”, Frege states that if we had a fictional thought about a mythical person whose real existence we later come to accept, “the thoughts would strictly remain the same” (1997d, 293), which means the sense is not, at least here, a question of existing in the world. And Frege absolutely does not require the whole domain (of quintillions of objects) to be grasped as a psychological act. Therefore, contrary to the allusion of the completeness fragment, senses of whole propositions should be graspable, even if they contain indeterminate concept-words.

Vague concept-words having a sense is more controversial. On one hand, Frege seems to say in “Concept and Object”, regarding the example of ‘*μωλν*’ discussed above, that it has a sense just as the fictional name

<sup>42</sup> Such examples originate with Esenin-Volpin (Wright 1996, 155).

<sup>43</sup> “My grasping” could be read as “my linguistic community’s grasping” and the argument would stay the same.

“Nausicäa” has one. Against that, Williamson argues that Frege’s theoretical apparatus cannot accommodate concept-words with sense but without reference. In the true proposition “There is a heap of sand on most building sites”, if concept-words contributed to the overall thought, the truth-value of the whole would be determined by finding the referent of those concept-words, among other conditions (Williamson 1994, 45). The way out is to observe that for Frege, the entire proposition cannot be true, since it contains concept-words without a referring concept. Thus, vague predicates could have senses that contribute to the *thoughts* of their containing propositions, as long as we’re prepared to read them similarly to fictional concepts. Unfortunately, there is no apparatus in Frege’s work to distinguish further the fictional from the indeterminate.

## 5. Conclusion

At the beginning of “Begriffsschrift”, Frege believed that natural language hides contents linked logically by pure thought but also that there exist “illusions that through the use of language often almost unavoidably arise concerning the relations of concepts’ (1997a, 51). While he uncovered many of those illusions by revealing a hidden logical structure beneath, Frege did not find a place for vague predicates, up to the point of taking the entire ordinary language as inconsistent on their account.

While some of Frege’s arguments seem stronger than others, some objectionable consequences include the following. First, Frege’s theory commits us to enigmatic objects such as the referent of “the concept ‘man’” or value-ranges. Secondly, he does not distinguish incompleteness of definition from vagueness as illustrated by the common contemporary issue of higher-order vagueness. Thirdly, we are seemingly forced to treat ordinary language, full of vague expressions as it is, as akin to fiction.

Finally, the researcher can invoke one of the competing contemporary theories of vagueness to paste vagueness onto Frege’s ontology, and I have illustrated briefly how epistemicism, fuzzy logic, trivaluationism, and supervaluationism can do that. The upshot is that vagueness does not raise, by itself, any important objection to Frege’s

project, even without going non-classical. Had he not treated heuristic issues as constitutive, he could accept that there are precise yet unwieldy definitions of 'Christian' to be discovered, that 'Heap' associate any individual group of sand to the True and the False – even though perhaps not by a simple gradual rule (Călborean 2020, 96) - and thus, that there is no parallel between fictional and vague predications. Capturing natural language in Frege's ontology is not impossible, it is only a little difficult.

### REFERENCES

- Beaney, Michael. "Introduction." In *The Frege Reader*, edited by Michael Beaney, 1–46. Wiley-Blackwell, 1997.
- Călborean, Marian. "The Nature and Logic of Vagueness." PhD Thesis, University of Bucharest, 2020.
- Dummett, Michael. "Wang's Paradox." In *Vagueness: A Reader*, edited by Rosanna Keefe and Peter Smith, 99–118. MIT press, 1996.
- Edgington, Dorothy. "Vagueness by Degrees." In *Vagueness: A Reader*, edited by Rosanna Keefe and Peter Smith, 294–316. MIT press, 1996.
- Fine, Kit. "Vagueness, Truth and Logic." *Synthese* 30 (3/4): 265–300, 1975.
- Frege, Gottlob. *The Foundations of Arithmetic*. New York: Harper & Brothers, 1960.
- Frege, Gottlob. "Begriffsschrift." In *The Frege Reader*, edited by Michael Beaney, 47–78. Wiley-Blackwell, 1997a.
- Frege, Gottlob. "[Comments on Sinn and Bedeutung]." In *The Frege Reader*, edited by Michael Beaney, 151–171. Wiley-Blackwell, 1997b.
- Frege, Gottlob. "Function and Concept." In *The Frege Reader*, edited by Michael Beaney, 130–48. Wiley-Blackwell, 1997c.
- Frege, Gottlob. "Introduction to Logic." In *The Frege Reader*, edited by Michael Beaney, 293–98. Wiley-Blackwell, 1997d.
- Frege, Gottlob. "Letter to Jourdain, Jan." In *The Frege Reader*, edited by Michael Beaney, 319–21. Wiley-Blackwell, 1997e.
- Frege, Gottlob. "Letter to Marty." In *The Frege Reader*, edited by Michael Beaney, 79–83. Wiley-Blackwell, 1997f.

- Frege, Gottlob. "Notes for Ludwig Darmstaedter." In *The Frege Reader*, edited by Michael Beaney, 362–67. Wiley-Blackwell, 1997g.
- Frege, Gottlob. "On Concept and Object." In *The Frege Reader*, edited by Michael Beaney, 181–93. Wiley-Blackwell, 1997h.
- Frege, Gottlob. "On Sinn and Bedeutung." In *The Frege Reader*, edited by Michael Beaney, 172–80. Wiley-Blackwell, 1997i.
- Frege, Gottlob. "Thought." In *The Frege Reader*, edited by Michael Beaney, 325–45. Wiley-Blackwell, 1997j.
- Frege, Gottlob. "Begriffsschrift, Eine Der Arithmetischen Nachgebildete Formelsprache Des Reinen Denkens." In *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, edited by Jean van Heijenoort. Harvard University Press, 2002.
- Frege, Gottlob. *Basic Laws of Arithmetic*. Oxford, New York: Oxford University Press, 2016.
- Heck, Richard G. *Reading Frege's Grundgesetze*. Oxford University Press, 2012.
- Heck, Richard G., and Robert May. "The Function Is Unsaturated." In *The Oxford Handbook of The History of Analytic Philosophy*, edited by Michael Beaney, 1st ed., 824–50. Oxford University Press, 2013.
- Heijenoort, Jean van. "Frege and Vagueness." In *Frege Synthesized*, edited by Leila Haaparanta and Jaakko Hintikka, 31–45. Dordrecht: Springer Netherlands, 1986. [https://doi.org/10.1007/978-94-009-4552-4\\_3](https://doi.org/10.1007/978-94-009-4552-4_3).
- Keefe, Rosanna, and Peter Smith. "Introduction: Theories of Vagueness." In *Vagueness: A Reader*, 1–57. MIT press, 1996a.
- Keefe, Rosanna, and Peter Smith. *Vagueness: A Reader*. MIT press, 1996b.
- Machina, Kenton. "Truth, Belief and Vagueness." In *Vagueness: A Reader*, edited by Rosanna Keefe and Peter Smith, 174–203. MIT press, 1996.
- Puryear, Stephen. "Frege on Vagueness and Ordinary Language." *The Philosophical Quarterly* 63 (250): 120–40, 2013. <https://doi.org/10.1111/j.1467-9213.2012.00103.x>.
- Tye, Michael. "Sorites Paradoxes and the Semantics of Vagueness." In *Vagueness: A Reader*, edited by Rosanna Keefe and Peter Smith, 281–93. MIT press, 1996.
- Unger, Peter. "There Are No Ordinary Things." In *Vagueness*, 3–40. Routledge, 2017.

- Varzi, Achille C. "Supervaluationism and Its Logics." *Mind* 116 (463): 633–76, 2007.
- Weiner, Joan. "Understanding Frege's Project." In *The Cambridge Companion to Frege*, edited by Thomas G. Ricketts and Michael Potter, 32–62. Cambridge University Press, 2010.
- Williamson, Timothy. *Vagueness*. London: Routledge, 1994.
- Williamson, Timothy. "Vagueness and Ignorance." In *Vagueness: A Reader*, edited by Rosanna Keefe and Peter Smith, 265–80. MIT press, 1996.
- Wright, Crispin. "Language-Mastery and the Sorites Paradox." In *Vagueness: A Reader*, edited by Rosanna Keefe and Peter Smith, 151–74. MIT press, 1996.



## DOES THE MEANING OF LYING POSE A PROBLEM TO PINOCCHIO'S PARADOX?<sup>1</sup>

MARIA-FLORIANA GAȚE<sup>2</sup>

**Abstract:** In this paper I sketch two solutions to Pinocchio's Paradox, mainly by resorting to the concept of lying, as it is conceived by Augustine in his "De mendacio". I will argue that the paradox is based on a slightly narrow conception of what it means to lie, as it confuses the meaning of lying with the meaning of falsity from the correspondence theory of truth. Furthermore, I will claim that the problems that I have highlighted are sufficient conditions to block the entire paradox and, therefore, to show that the argument that supposedly leads to a contradiction is actually fallacious.

**Keywords:** Pinocchio's paradox, truth, falsehood, lie, Augustine, intentionality

### 1. Introduction

Paradoxes have always played an important role in the history of philosophical thought, being "associated with crisis in thought and revolutionary advances" (Sainsbury 2009, p. 1) because its main effect is

---

<sup>1</sup> This paper is an improved version of a presentation I held during the Online National Conference of Theoretical Philosophy for Students, which is virtually available here: <https://filosofieteoretica.wordpress.com/2022/04/11/maria-floriana-gate-ce-inseamna-sa-minti-o-propunere-de-blocare-a-paradoxului-lui-pinocchio/>

<sup>2</sup> M.A. student at the Faculty of Philosophy, University of Bucharest. Contact: [maria.gate@s.unibuc.ro](mailto:maria.gate@s.unibuc.ro).

to determine “major reconstruction at the foundations of thought” (Quine 1966, p. 1), especially starting from where it leads to an obvious contradiction. Perhaps one of the most famous and intuitive paradoxes is the classical version of the Liar (also called simple-falsity Liar): “What I am now saying is false.”, forms of which date back to the paradox of Eubulides (Eubulides the Cretan says “All Cretans are liars.”).

Some classical solutions to this paradox are Tarski’s (1944) or Kripke’s (1975) well-known semantic hierarchies, but we can imagine one scenario in which these theories hold no value: what if Pinocchio, whose nose grows if and only if it tells a lie, says “My nose is growing.”? The last decade of philosophical debates regarding Peter Eldridge-Smith’s (2010) presumably non-semantic version of the Liar paradox has proven to be very prolific in terms of researching the assumptions it is based on or rather exploring the metaphysical, logical and epistemological consequences that come with its acceptance.<sup>3</sup>

In the following sections I will first highlight the main characteristics of a paradox, so that we can later establish the *Pinocchio scenario* with all its specific features as being paradoxical. Secondly, I will sketch a solution to this paradox by resorting to Augustine’s conception on lying and arguing that the paradox is based on a questionable assumption that considers falsehood to be synonymous with lying.

Thirdly, I will propose another way of resolving the paradox that has a connection with the first because it questions the nature of Pinocchio and its compatibility with our definition of what lying is. Lastly, I will tackle some possible counterarguments that may arise following my arguments, precisely regarding the two solutions and the claim that they may not take into consideration the fact that the Pinocchio scenario is a fictional one and there may not be any useful insights resulting from trying to solve it with a contextualist solution.

---

<sup>3</sup> For reference, see Beall (2012, 2014), Eldridge-Smith (2011, 2012), D’Agostini and Ficara (2020) etc.



## 2. What, exactly, is a paradox?<sup>4</sup>

Given the fact that this paper's aim is to propose two ways of resolving Pinocchio's paradox, I find it necessary to first clarify what is meant by the term "paradox" and what exactly means to resolve one, if possible. In doing so, I will mainly follow W.V.O. Quine's classical understanding of what is a paradox from *The Ways of Paradox* (1966) and I will later argue that the *Pinocchio scenario* as proposed by Eldridge-Smith (2010) is indeed paradoxical even though certain conditions need to be fulfilled - and it belongs to the class of paradoxes that can be solved by proving the reasoning that leads to it is, in some way, fallacious.

Most often, a paradox is generally understood as "an apparently successful argument having as its conclusion a statement or proposition that seems obviously false or absurd" (Lycan 2010, p. 615). Sainsbury also commits to a similar definition of a paradox: "an apparently unacceptable conclusion derived by apparently acceptable reasoning from apparently acceptable premises" (Sainsbury 2009, p. 1). Even though Quine admits that by considering any such argument followed by an absurd conclusion a paradox many things may be "left unsaid" (Quine 1966, p. 1), I think this account still is very much relevant for my present aim.

Following Quine's definition we can distinguish two possible ways of resolving paradoxes: either we can "show [...] that (and why) despite appearances the conclusion is true after all" (Lycan 2010, p. 615) or we can show that "the argument is fallacious" (Lycan 2010, p. 616). The latter case, which will be relevant for my present aim, implies that there must be an error that must come from the interior of a premise or even from the reasoning that leads to the contradictory (paradoxical) conclusion itself.

Quine proposes a taxonomy of paradoxes that implies three distinct types or groups: veridical paradoxes, falsidical paradoxes and antinomies. Among the paradoxes of the first type we can find examples such as Russell's famous Barber paradox or Frederic from *The Pirates of Penzance* (Quine 1966, p. 3). Veridical paradoxes are the ones whose "conclusion is in fact true despite its air of absurdity" (Lycan 2010, p. 616).

---

<sup>4</sup> The title of this chapter is also shared by a paper written by William G. Lycan (2010)

On the other hand, falsidical paradoxes are the ones whose conclusion is “obviously false or self-contradictory” (Lycan 2010, p. 616) and some errors from the specific proof can be detected as responsible for the fallacious argument. One example of such a paradox is De Morgan’s proof that  $2=1$  or, as I will show, even Pinocchio’s paradox.

Lastly, the third type of paradoxes, also called antinomies, are the ones that cannot be traced back and, therefore, cannot be resolved in the two aforementioned ways. These are the ones that pose a problem to human thought and can make us reconsider the very foundations of our thinking. One such example is, according to Quine, Grelling’s paradox regarding the “heterological, or non-self-descriptive, adjectives” (Quine 1966, p. 4).

### 3. Pinocchio's paradox

In his article, Peter Eldridge-Smith offers a formulation of Pinocchio’s Paradox that is as follows:

“Pinocchio’s nose grows if and only if (iff) what he is stating is false, and Pinocchio says ‘My nose is growing’. So, Pinocchio’s nose is growing iff it is not growing.” (Eldridge-Smith & Eldridge-Smith 2010, p. 213)

According to my reading of this aforementioned statement, the reasoning can be succinctly reformulated in three distinct sentences that form the paradoxical argument, such as follows:

- (1) Pinocchio utters “My nose is growing”.
- (2) Pinocchio’s nose grows if and only if (iff) he tells a lie.
- (3) Pinocchio’s nose grows iff it does not.

Here it can be observed the classical structure of a paradox: first, we have proposition (1), which I will call the *Pinocchio statement*, then we have proposition (2) consisting of a biconditional that makes the entire displayed scenario (possibly) paradoxical. Finally, proposition (3) is a

contradiction that is easily obtained by substituting Pinocchio's utterance from (1) in the right side of the biconditional from (2): Pinocchio's nose grows iff his previous statement is a lie, thus his nose is not growing.

An important modification that I have made to the original structure of the paradox is that I have replaced the *Pinocchio condition* with a more accurate version according to the fictional story of Pinocchio as created by Carlo Collodi: "Pinocchio's nose grows if and only if what Pinocchio is saying is not true." (Eldridge-Smith 2012, p. 751) has been replaced with "Pinocchio's nose grows if and only if he tells a lie."

Eldridge-Smith considers this paradox to be a version of the Liar paradox, but one that has the advantage of not using a semantic predicate. He considers that the predicate "is growing" is rather empirical than semantic and is only used to express a "facial [...] feature" (Eldridge-Smith 2010, p. 213). Even though it can be argued that the causal chain of this predicate actually leads to the use of truth-functions, this will not be the objective of this paper.

I consider this argument to be part of the falsidical paradoxes group as described by Quine mainly because it seems to have two core assumptions, even though tacitly presumed non-problematic, that give a false appearance of argumentative validity. I will sketch the core ideas of these assumptions as follows:

(A1) The mismatch between language and reality (from now on I shall refer to this as a "falsehood") is synonymous with the mismatch between a speaker's thoughts and the language she uses to express them (called a "lie").

(A2) Pinocchio's nature is analogous to a machine that processes Pinocchio's utterance and turns it into nose growth if and only if that utterance is a lie.

For the particular examination of these problematic assumptions I will reserve the following two sections and I will consider them as constituting the main objects of a possible refutation of this argument's premises.

#### 4. Augustine's *De mendacio* and the first way of resolving the Pinocchio's paradox

In this section I aim to sketch a solution to Pinocchio's paradox that relies on Augustine's view on what lying is. In order to do that, I will argue that the second premise of the previous stated argument is based on a very narrow conception of what lying is, that identifies it with falsehood.

The second premise, the *Pinocchio condition*, is a biconditional that directly links Pinocchio's nose growth with his utterance of a lie. However, a problem arises in the moment we wish to do the substitution: what reasons do we have in order to think that there is a connection between - almost like a "causal [...] relation" (Eldridge-Smith 2010, p. 213) - the utterance of a lie and the ontological consequence of a growing nose? My thesis is that this is only possible if we tacitly assume the correspondence theory of truth as providing a framework for the ascription of the truth-value of Pinocchio's utterance. In his *Metaphysics*, Aristotle offers the classical formulation<sup>5</sup> of the correspondence theory of truth, making it clear that the truth-value of a proposition stands in its accordance with the factual reality - thus introducing an important ontological commitment to "an ontology of facts (or states of affairs)" (Tomi 2020, p. 116).

The definition of lying has been expressed in various different forms throughout the history of philosophy, but definitely one account still stands out among the others, namely Augustine's. His first treatise on mendacity, *De mendacio*, is "one of the first attempts in Western scholarship to provide a systematic study of lying and to provide a concise, clear-cut and reliable definition of what constitutes a lie" (Gramigna 2013, p. 447). That being said, intentionality plays a key-role in Augustine's conception mainly because he seems to assume that lying is a phenomenon that is bound to human beings<sup>6</sup>. Moreover, the intention

---

<sup>5</sup> "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, it is true." (Aristotle, *Metaphysics* 1011b25)

<sup>6</sup> Even though this may be a consequence of Augustine's Christian view on the immorality of lying, I will assume that this consideration does not directly influence

to deceive is not the only condition that an utterance needs to fulfill in order to be considered a lie: duplicity in thought (lat. *duplex cogitatio*) implies that the liar has mismatching “inwardly concealed believed truth and outwardly expressed truth” (Gramigna 2013, p. 452).

Augustine defines the act of lying such as follows, only resorting to the use of the linguistic and mental planes:

“Wherefore, that man lies, who has one thing in his mind and utters another in words, or by signs of whatever kind. Whence also the heart of him who lies is said to be double; that is, there is a double thought: the one, of that thing which he either knows or thinks to be true and does not produce; the other, of that thing which he produces instead thereof, knowing or thinking it to be false.” (Augustine, *De mendacio* III.3.)

One important consequence of this previous definition is that it implies a certain view on the nature of a lie - lifting up ‘the discussion of lying from an ontological paradigm [...] that tackled the problem of lying within the frame of objective falsehoods to the metalevel of the interpreter’s beliefs and intentionality’ (Gramigna 2013, p. 449):

“For from the sense of his own mind, not from the verity or falsity of the things themselves, is he to be judged to lie or not to lie.” (Augustine, *De mendacio* III.3.)

“Whence it comes to pass, that he may say a false thing and yet not lie, if he thinks it to be so as he says although it be not so; and, that he may say a true thing, and yet lie, if he thinks it to be false and utters it for true, although in reality it be so as he utters it.” (Augustine, *De mendacio* III.3.)

This view on what lying is strongly connected to Augustine’s view on the purpose of language and how it should be. Similarly, Quine seems to assume that what we express by means of language have to be in accordance with logical principles (such as the principle of non-contradiction) and that paradoxes are just problematic cases – “crises in

---

any other aspect that may be part of his philosophy of language and epistemology regarding the concept of lying

thought” - that need to be solved. In other words, language should be in accordance with logic, not the other way around. These two views can be associated with Augustine's theory of the main functions of language from *De magistro* and it can be said that “liars contradict [the very principle of communication - the transmission of knowledge]” (Gramigna 2013, p. 456).

### 5. Second way of resolving the paradox

This second solution to Pinocchio's paradox is based on the same premise as the first solution, namely premise (2), the *Pinocchio condition*, albeit approaches from a different perspective: how can be Pinocchio's nose growth (directly) linked to telling a lie without “deliberate conscience, on the part of the speaker” (D'Agostini & Ficara 2020, p. 252). This way of resolving the paradox resides in questioning the implied nature of Pinocchio that seems to be almost similar to a machine based on a very rigid algorithm (once it tells a lie, its nose is automatically growing) - or a “truth barometer” (Eldridge-Smith 2012, p. 750), as it has been called before. The important question is – what reason have we to believe that such a thing could possibly exist? Or more precise – even though its existence is possible, is it also plausible?

This objection, also called ‘the objection of fictionality’ (D'Agostini & Ficara 2020, p. 252), has been first raised by JC Beall (Beall 2011) and then tackled by D'Agostini and Ficara (D'Agostini & Ficara 2020), ultimately resulting in proposing the Blushing Liar paradox, which is a version of the Liar paradox similar to Pinocchio's paradox that does not have a problem with plausibility. These paradoxes have the same structure, since they both contain a character that utters a self-referencing remark about their physical characteristics, that are non-semantic (the blushing of the cheeks or the growing of the nose).

My thesis is that this condition on which the paradox takes place seems to be in direct contradiction with multiple instances in which Eldridge-Smith has described Pinocchio as acting similarly to a conscious agent – “Pinocchio was beguiled” (Eldridge-Smith 2012, p. 749) etc. The

key-question is: "What exactly is the nature of Pinocchio?". If Pinocchio is just a machine, a "truth barometer", the Pinocchio scenario is not paradoxical because, as I have argued before, he cannot possibly be capable of lying because he must be capable of deliberately and consciously telling a lie. Also, telling a lie in the Augustinian view implies the intention of the speaker and also an incongruence between what he thinks to be true and what he claims to be true.

## 6. Possible counterarguments

One counterargument that has been used against the Pinocchio's Paradox is a reversed form of JC Beall's objection of fictionality that says that, because my previous argumentation seems to focus on the fact that Pinocchio does not have the characteristics of a real, actual, world person, he cannot possibly utter a statement that represents a lie. This objection appeals to the use of a modal apparatus and argues that fictionality is not the same thing as possibility and, thus, cannot imply further ideas about the actual world. However, it could still be argued that the two solutions that I have proposed can be subjected to the same type of critique: what if the *Pinocchio scenario* is replaced with a more plausible scenario in our actual world, such as the heart rate of someone that says "My heart rate goes up right now."<sup>7</sup>

My response to this counterargument, however, implies making a clear distinction between the substrate ideas of my argumentation and the assumptions of this type of critique. First of all, the two solutions that I have proposed do not imply or necessitate the idea of availability only in the actual world. Rather, they only imply that the idea of lying is very

---

<sup>7</sup> This is, according to my interpretation, a reformulated version of the Blushing Liar paradox, since it has the same structure as Pinocchio's paradox, but it does not have a problem with plausibility. I would like to address my thanks and accredit this formulation to the anonymous reviewer that read my text and suggested this example as a possible counterargument to my perspective

different from falsehood and this consideration has no modal commitments. Of course there can be cases in which falsehoods coincide with lies (and they are, perhaps, even the majority), but it is necessary to take a close look at the exceptions and to acknowledge the fact that a paradox cannot be based on a relation that is not identity between them.

A common use in the literature of specialty for the objection of fictionality is to block the claim of Pinocchio's Paradox to have implied the confirmation of metaphysical dialetheias (D'Agostini & Ficara 2020, p. 256) because of the fact that fictionality is not the same possibility. Moreover, this counterargument can also be extended to the idea that, due to the fictional character of the *Pinocchio scenario*, it cannot possibly represent a paradox because it does not satisfy Quine's first condition of a paradox. However, the fictional character seems to be one of the additional details about the *Pinocchio scenario* that don't actually have direct consequences upon the idea that there might be something more to the concept of lying than the non-correspondence between someone's utterance and the state of affairs.

Instead of this "fictional *vs.* actual approach", my vision implies a much broader, contextualist view upon core, essential, concepts from the specialty literature discussions about Liar-like paradoxes. In this sense, I think there is much more to learn from Pinocchio's paradox besides exploring its consequences in the framework of assuming its exclusively fictional character. Thus, my general approach and response to these types of critiques implies the assumption of a contextualist point of view upon paradoxes in general, and Pinocchio's paradox in particular. To support my previous claims, I very much adhere to Michael Glanzberg's (2001) general contextualist considerations, such as follows:

"[...] paradoxes can be much more. Beyond posing some logical puzzle, they can indicate deep problems of some kind. A solution to a paradox of this sort involves more than just finding an appropriate logical trick. It requires identifying the source of some apparent inconsistency, and explaining why it is merely apparent." (Glanzberg 2001, pp. 217–218)



## 7. Conclusion

In conclusion, in this paper I have proposed two connected ways of resolving the Pinocchio's paradox by exposing its deceiving appearances and arguing that it is, in fact, a falsidical paradox in Quine's sense - a fallacious argument based on traceable errors. The reasoning behind this assertion can be briefly reconstructed in the following argument:

- (1) Pinocchio utters "My nose is growing".
- (2) Pinocchio's nose grows if and only if he tells a lie.
- (3) A lie is not the same thing as a falsehood.

---

(C1) Telling a factual falsehood does not exclude the possibility of Pinocchio's nose growing.

- (4) Telling a lie implies the intention to deceive and a mismatch between the speaker's thoughts and language.

---

(C2) If Pinocchio's nose grows, he must have had the intention to deceive and a mismatch between the speaker's thoughts and language implies that the possessor must be a conscious agent.

- (5) The possession of the intention to deceive and a mismatch between the speaker's thoughts and language implies that the possessor must be a conscious agent.

---

(C3) Pinocchio must be a conscious agent, capable of making deliberate decisions.

But, at last, (C3) is not compatible with (2) because there cannot be imposed such a condition that implies a direct relation between a deliberate act, *i.e.* lying, and an empiric feature. I think that these conceptual errors included in Pinocchio's paradox can teach us many valuable lessons about the way we conceptualize falsehood and lying or, probably most important, the dependence of some ideas to human particularities - such as awareness, agency etc.

## REFERENCES

- Aristotle (1928). *Metaphysics. The Works of Aristotle*, Vol. VIII, Oxford: Clarendon Press.
- Augustin (1995). *De magistro. Ediție bilingvă*, București: Ed. Humanitas.
- Augustin (2016). *Despre minciună. Ediție bilingvă*, București: Ed. Humanitas.
- Beall, JC (2012). "Dialetheists against Pinocchio". *Analysis*, Vol. 71, No. 4, October 2011, pp. 689-691.
- Beall, JC (2014). "Rapunzel shaves Pinocchio's Beard". *Contradictions. Logic, History, Actuality*, Berlin: Walter De Gruyter GmbH, pp. 27-30.
- Carson, T. L. (2006). "The Definition of Lying". *Nous* 40:2, pp. 284-306.
- Chisholm, R. M. & Feehan, T. D. (1977). "The Intent to Deceive". *Journal of Philosophy*, No. 74 (3), pp. 143-159.
- Collodi, C. (1916). *Pinocchio. The tale of a puppet*, Racine: Whitman Publishing Co.
- D'Agostini, F. & Ficara, E. (2020). "The Blushing Liar", *Philosophia: International Journal of Philosophy*, Vol. 21, No. 2, June 2020.
- Eldridge-Smith, P. & Eldridge-Smith, V. (2010). "The Pinocchio Paradox". *Analysis*, Vol. 70, No. 2, April 2010, pp. 212-215.
- Eldridge-Smith, P. (2011). "Pinocchio against the dialetheists". *Analysis*, Vol. 70, No. 2, April 2011, pp. 306-308.
- Eldridge-Smith, P. (2012). "Pinocchio beards the Barber". *Analysis*, Vol. 72, No. 4, October 2012, pp. 749-752.
- Glanzberg, M. (2001). "The Liar in Context". *Philosophical Studies*, Vol. 103, pp. 217-251.
- Gramigna, R. (2013). "Augustine on lying: A theoretical framework for the study of types of falsehood", *Sign Systems Studies* 41 (4), pp. 446-487.
- Kripke, S. (1975). "Outline of A Theory of Truth". *The Journal of Philosophy*, Vol. 72 (19), pp. 690-716.
- Lycan, W. G. (2010). "What, exactly, is a paradox?". *Analysis*, Vol. 70, No. 4, October 2010, pp. 615-622.
- Quine, W.V.O. (1966). "The Ways of Paradox". *The Ways of Paradox and Other Essays. Revised and enlarged edition*, Cambridge: Harvard University Press.

- Sainsbury, R.M. (2009). *Paradoxes. Third Edition*, New York: Cambridge University Press.
- Tarski, A. (1944). "The Semantic Conception of Truth and the Foundations of Semantics". *Philosophy and Phenomenological Research*, Vol. 4, No. 3 (March 1944), pp. 341-354
- Tomi, P. (2020). "Few Critiques on the Classical Truth Theories", *Annals Of The University Of Bucharest – Philosophy Series*, 69 (2), 115-124.

*Reproducerea integrală sau parțială, multiplicarea prin orice mijloace și sub orice formă, cum ar fi xeroxarea, scanarea, transpunerea în format electronic sau audio, punerea la dispoziția publică, inclusiv prin internet sau prin rețele de calculatoare, stocarea permanentă sau temporară pe dispozitive sau sisteme cu posibilitatea recuperării informațiilor, cu scop comercial sau gratuit, precum și alte fapte similare săvârșite fără permisiunea scrisă a deținătorului copyright-ului reprezintă o încălcare a legislației cu privire la protecția proprietății intelectuale și se pedepsesc penal și/sau civil în conformitate cu legile în vigoare.*

*Acts such as the permanent or temporary reproduction by any means and in any form, in part or in whole, or acts such as photocopying, scanning, conversion into electronic or audio format, public distribution over any network, permanent or temporary storage on devices or systems with the possibility of recovering information, for commercial or free purposes, are expressly prohibited without prior written permission of the copyright holder. Committing such acts without the written permission of the copyright holder, represents a violation of the legislation regarding the protection of intellectual property, copyright violations are subject to civil and/or criminal sanctions.*

---

tipografia.unibuc@unibuc.ro

Bd. Iuliu Maniu 1-3, Complex LEU  
tel: 0799 210 566

Tiparul s-a executat la Tipografia  
**Editurii Universității din București – Bucharest University Press (EUB – BUP)**

---