SYNONYMY AND IDENTITY OF PROOFS

A Philosophical Essay

Dissertation
zur
Erlangung des akademischen Grades
Doktor der Philosophie
in der Philosophischen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von
Tiago Rezende de Castro Alves

aus

Niterói, Brasilien

2019

Gedruckt mit Genehmigung der Philosophischen Fakultät
der Eberhard Karls Universität Tübingen

Dekan: Prof. Dr. Jürgen Leonhardt

Hauptberichterstatter: Prof. Dr. Peter Schroeder-Heister

Mitberichterstatter: Prof. Dr. Luiz Carlos P. D. Pereira
Prof. Dr. Thomas B. Sattig

Tag der mündlichen Prüfung: 30. Juli 2018

Universitätsbibliothek Tübingen TOBIAS-lib

To the memory of Maria Rosa dos Santos
and to all the people upon whose pain and sorrow
life is a feast.

ACKNOWLEDGEMENTS

I am most grateful to Prof. Peter Schroeder-Heister for all his many qualities as a supervisor, but most especially for the generosity and unwavering respect he showed towards myself and my work from the very beginning of my doctoral studies.

I would also like to thank Prof. Luiz Carlos Pereira for all his invaluable teachings, as well as for his kindness and support throughout my whole academic trajectory. I owe him more than words could ever express.

I am indebted to Germany and its people for having awoken me from a comfortable lethargy and forced me to acknowledge how urgently we should change a great many crucial aspects of our life.

Finally, my deepest and warmest thanks to Luana de Castro Alves, my beloved wife. Much of the essence of this work is the result of the influence of some of her ideas upon me. Her sheer brilliance as a thinker, a lover and a person is to me a source of unending inspiration and fuels my will to become a better person every single day of our lives. Thank you for existing.

CONTENTS

“Forgive me this my virtue,

For in the fatness of these pursy times

Virtue itself of vice must pardon beg –

Yea, curb and woo for leave to do him good”

Hamlet , Act 3, Scene 4 - W. Shakespeare.

Preface

If the 21st century were a day, the clock would now be at around 4:30 in the morning. While this text is being written, we are, in many senses, witnessing the verge of its dawn. And in using this phrase as so many before myself already have, I would like to, for once, reverse the order of the emphasis usually put upon its parts. Rather than saying that we are witnessing *the verge of its dawn,* I would like to stress that we are *witnessing* the verge of its dawn. And we see nothing. We look at the horizon with our eyes wide open, waiting for the sun to emerge and allow us to see the land before us with its generous light, so we can take a step that will lead us safely to where we want. But it is still too dark – and while we still cannot see the intricacies and obstacles that await on the path that lies ahead of us, we decide to stand still, listening to the clock ticking in quiet anxiety and entertaining thoughts far darker than our surroundings about what we will see when the sun finally rises.

The atmosphere of our day – mimicked to the best of my ability in the former paragraph – insistently suggests that a certain aura of tragedy and respectability emanates from this picture. We feel that our hands are tied by our unavoidable powerlessness over what lies in the oblivion, and this makes the quietly desperate wait for the sunrise feel not only inescapable, but also mandatory, and even further, right and justified. Many agree that taking paths we have already been through – paths that we could thus easily follow from memory, light and even eyes being totally dispensable – is not an option, for we know they lead to no place to where we are willing to go. Some of these explicitly endorse the stillness in which they see themselves, condemning every sketch of a move as reckless and threatening; while others urge that such stillness cannot go on, only to face what they perceive as an even harsher desperation provoked by the immobilising effect that their lack of vision exerts upon them. There are also those who perceive the stillness imposed by this time as a straightjacket, or their own live burial, and do not care to avoid going along old familiar roads, as long as they are allowed to move freely again. And all fail to have the capability to even conceive of a fourth option as a real possibility.

From the point where I now stand, this is nevertheless nothing but a poor taste, egocentric and self-apologetic melodrama told by their own characters to themselves, and which keeps them entertained enough not to be susceptible to listen to anyone or anything else. This work is thus an attempt at giving these characters something else to which to listen.

The rather tense historical period which we now live can be described in many ways; but a way of grasping this “post-truth” *Zeitgeist* in a fruitful way is to take it as a radical contestation of the very notion of knowledge – or any contextually adapted guise of it – as touchstone of a possible criterion to help us define or pursue legitimate ethical-political goals. It is a far more radical situation than the epistemic-foundational crisis of the 19th century; for it is not a crisis caused by the lack of criteria to fill the vacancy of the formal *loci* of knowledge and foundation, but rather by the gradual implosion, the ongoing conceptual shrinkage of these very formal *loci* towards senselessness and extinction, to which a series of historical developments has been leading. Plato’s paradise is too small; and unless we are somehow cut to fit in it, it shall not give us any shelter.

In early days, when West was still less than the vaguest plot of a nightmare, Aristotle opened the doors to the formal investigation of language with aims which belonged in his time; a time when knowledge was, from the political viewpoint upon which its sense unavoidably depends, still nothing but a promising possibility – one which he, following Plato and preceding almost all subsequent heroes of our tradition, vigorously embraced. These aims led him to forge, in the *Prior Analytics*, the notion which is now to be the object of our scrutiny: *συλλογισμός*; or, as the literary tradition from which this work descends calls it, proof. This notion, as well as its investigation, was envisaged and concretised as a means to the construction of the ideal of knowledge, the living seed of which was planted in fertile soil. The plant germinated, nurtured voraciously from it and grew up into a colossal tree, that, eventually, drained its very source of life; and it is now dead, even though it stands, imposing and stolid as ever, still firmly attached to the now dry and sterile soil from which it once emerged. The notion of proof, however, outlived it. In the absence of its original guiding ideal, though, what is to become of it? – or, better said, what are *we* to *make* of it?

T.R.C.A., Tübingen, October 2017

Prologue: a pseudo-historical account of the notion of proof as the conspicuous formal contours of moral apologies for claims to knowledge

Thales of Miletus is commonly regarded as the first philosopher due to the fact that, by means of his doctrine that “all is water”, he has arguably inaugurated a way of explaining the world; one that essentially consists of a radical reduction of the plurality of things and phenomena in general to an originary principle (*ἀρχή*). It is nevertheless a later milesian thinker, Anaximander, who is more likely responsible for bringing the business of *justification* to the heart of what would become the philosophical activity. The only surviving fragment of Anaximander’s own text was bequeathed by means of a quotation made by Simplicius:

“ἐξ ὧν δὲ ἡ γένεσίς ἐστι τοῖς οὖσι, καὶ τὴν φθορὰν εἰς ταῦτα γίνεσθαι κατὰ τὸ χρεών· διδόναι γὰρ αὐτὰ δίκην καὶ τίσιν ἀλλήλοις τῆς ἀδικίας κατὰ τὴν τοῦ χρόνου τάξιν” (In a free translation: “Whence origin is to things, thereinto is also perishment, according to necessity. For they concede to each other justice and reparation for the injustice in accordance with the ordinance of time.”)

Tradition says that Anaximander deemed *τὸ ἄπειρον* (“the undetermined” or “the boundless”) to be the *ἀρχή*, and this rather enigmatic fragment can be interpreted so as to help us make an interesting and illuminating picture of what he might have meant by that.

The realm of inquiry of so-called pre-socratic thinkers was *φύσις,* i.e. the nature – for lack of a better expression – as *becoming,* a pervasive and relentless flux in which and by force of which all things come into being and cease to be. It seems that Thales looked upon *φύσις* as being fundamentally a multiplicity of things, and in operating his radical and general reduction of it to water, viz. a single element of which everything else is not but a variation, he succeeded in giving an explanation to the *cohesion* of things – *ἀρχή* is thus given the role of first and foremost a principle of unity.

The fragment transcribed above seems to suggest that Anaximander, on the other hand, has been impressed not so much by the multiplicity of things that come about within the relentless flux of *φύσις*, but rather by the very relentlessness of the flux itself, of which the multiplicity that caught Thales’s attention is not but a symptom. The problem he has in hands is thus of a completely different nature; he feels the urge to account not for the compatibility between two apparently disparate aspects of *φύσις* – namely, the multiplicity of things and the cohesion of their totality – but rather for the *reason why,* in the first place, things inexorably come into being and, in so doing, inexorably cease to be – if there is any. The task given by Anaximander to his *ἀρχή* is thus to *justify* the very existence, the very there-being and persistence of *φύσις.*

According to the famous interpretation of Nietzsche, Anaximander's fragment suggests that he saw the coming to be of everything within *φύσις* essentially as "injustice" (*ἀδικία*) –, an unavoidably unexplainable peremptoriness. Notwithstanding, the Milesian saw in the perishment and decay that, according to a "monstruous experimental evidence" given by *φύσις* itself, inexorably assails all things that once flourish within the flux, a reparation for their "crime" of violating what would be the perfect balance of nothingness with their becoming. Ceasing to be is the price, the penalty to pay for coming into being – and so the relentless flux of *φύσις* provides itself with justification in a most concrete and literal way, as the expiation of its own unjustifiable coming into being.

But "whence" and "whereto"?; what could be the originary and explanatory principle of such perennial self-expiatory crime?; or, in other words, how is this tragic dynamic balance possible? Once the question is put this way, it is easier to make sense of *ἄπειρον*. Since any determination of a thing can only come about a result of and within the flux of *φύσις*, then nothing the being of which consists of having this or that determination may provide *φύσις* with the ontological "fuel" and sustentation it needs to justify itself – were it otherwise, then we would be somehow incurring in circularity. Hence water, earth, air, fire and in general anything whatsoever that can be positively defined by having given properties or determinations must be rejected as *ἀρχή* according to Anaximander's approach to *φύσις.* Indeed, it is easy to see that whatever may justify the persistence and there-being of *φύσις* must be essentially devoid of determinations; for only so it is free from the necessity of coming into being and perishing within the flux.

By creating the notion of the undetermined – *τὸ ἄπειρον –,* then, something which is by definition safeguarded from the vicissitudes of the flux, Anaximander succeeded in explaining the essential trait of a principle that justifies *φύσις* inasmuch as it alone can be the primal source, originary nature and ultimate destiny of everything that comes into being. By postulating its existence as the *ἀρχή*, the Milesian thinker provided not only a principle of cohesion, a common measure and originary matter to all things, but also a *justification* to the very there-being of *φύσις, –*i.e. an explanation of it which makes it clear not only that it has a cohesive and *internally* coherent structure, as Thales's does, but also that it respects a *demand* for cosmological *balance* which is in force independently of its being and structure.

Thus, in particular, Anaximander inescapably inserted deep in the core of the philosophical thinking what one could call its *tragic* element – which could be roughly described as the tendency to see and explain whatever happens to be the case as inexorable, inescapable, as somehow *necessary*. His reaction before such an awe-inspiring picture of the world was nevertheless quite unique: instead of falling on his knees in some form of despair or adoration, instead of fighting against or of letting himself innebriate and be taken along by the force of such an overwhelming flow, Anaximander's grasp of *φύσις* as impregnated with an essentially *moral* character took him to *question* the object of his inquiry: "Why is it so and not some other way?; why is it at all, in the first place?" Anaximander pioneered the attitude of *questioning the legitimacy* of that which is not justified, and in so doing opened the doors to what tradition would later call "burden of proof" – an originarily and essentially *moral* burden, put by him from the very outset of his inquiry upon the shoulders of anything that dares laying claim to being (the case) – in short, a demand for *justification*. It is thus by virtue of the sublime weight of the moral burden of justifying a still unjustified world that explanations were shaped into justifications in the very dawn of Philosophy – and so the *ἄπειρον* appeared as, in many and crucial senses, the key element of the first known non-mythical and non-alegorical explanation in our tradition of why things *must* be the way they are claimed to be.

Now, Anaximander's own way of questioning the legitimacy of his object of inquiry could be described as moderate; or, put another way, his attitude towards the flux of becoming could be quite understandably described as more apologetic than, say, challenging. For although he does indeed set himself in a quest for a justification of the flux – i.e. to what he experiences as the nature of *φύσις* –, he, as already suggested above, seems never to consider the possibility of actually *denying* (the existence of) it on any basis. But what if his quest turned out as a failure? Would he keep on accepting the there-being of an injust and unjustified reality merely by the force of its own persistent and vociferous self-affirmation?

Indeed, it seems that this vociferous self-affirmation was enough at least for Anaximander to suppose the existence of the flux beforehand and so, in a way or another, safeguard its ontological status from the potentially destructive outcome that an eventual failure to meet the demands of his questioning could have. The milesian has thus attempted to justify his depiction of reality as a means to, I repeat, free it from a *moral* burden; a burden it should carry even in case one should already feel compelled to accept its existence by virtue of other factors, which, it seems, he did.

It took a severe radicalisation of Anaximander's questioning attitude – or, put another way, an understanding of the meaning of this questioning as a *challenge* with respect to his object of inquiry – to add *existential* import to its originarily moral character. Such a radicalisation can be variously witnessed in some key episodes of later pre-socratic thought, having yielded quite different, even diametrically opposed philosophies. For instance: Heraclitus of Ephesus, on the one hand, denies the legitimacy of an eventual existential challenge to *φύσις* by denying the supposition that becoming is in any sense in need of (or even compatible with) a justification of any sort. This thinker observes that becoming is all there is, the supposition of there being anything alien to it being unavoidably groundless; it is thus impossible to justify it by appeal to anything other than itself. The observance of this total, absolute character of the flux reveals that its nature and there-being are themselves neither unjustified nor justified; for justice, injustice, well-groundedness and indeed all things are possible only as unfoldings within the flux. Therefore, challenging becoming and its there-being on the basis of its supposed lack of justification is not only failing to understand its total, pervasive and necessarily self-sustaining character, but also its intrinsic incompatibility with the sort of moral questioning posed by Anaximander. The vociferous self-affirmation of relentless becoming referred before is completely innocent, inasmuch as it is essentially *amoral*, and it is senseless not only to condemn it, but also to deny it in any sense; for such self-affirmation by its very nature *defines* what it means to exist. Nietzsche provides us with an image, as usual very eloquent, that might help one understand this:

"Ein Werden und Vergehen, ein Bauen und Zerstören ohne jede moralische Zurechnung in ewig gleicher Unschuld hat in dieser Welt allein das Spiel des Künstlers und des Kindes. Und so, wie das Kind und der Künstler spielt, spielt das ewig lebendige Feuer, baut auf und zerstört, in Unschuld – und dieses Spiel spielt der Äon mit sich. Sich verwandelnd in Wasser und Erde, türmt er wie ein Kind Sandhaufen am Meere, türmt auf und zertrümmert: von Zeit zu Zeit fängt er das Spiel von neuem an. Ein Augenblick der Sättigung: dann ergreift ihn von neuem das Bedürfnis, wie den Künstler zum Schaffen das Bedürfnis zwingt. Nicht Frevelmut, sondern der immer neu erwachende Spieltrieb ruft andre Welten ins Leben. Das Kind wirft einmal das Spielzeug weg: bald aber fängt es wieder an in unschuldiger Laune. Sobald es aber baut, knüpft, fügt und formt es gesetzmäßig und nach inneren Ordnungen."[1]

In this fashion, Heraclitus depicts becoming – even more tragically – as essentially spontaneous, innocent, and ontologically self-sufficient, thus being both existentially and morally unchallengeable; and hence free from the burden of justification and necessarily inexorable from the very outset. The radicalisation of the understanding of the meaning of Anaximander's demand for justification with respect to its object in this case lies in the following: instead of considering that *φύσις* answers (or fails to do so) in any sense to Anaximander's demand for justification, Heraclitus deems such demand as plainly incompatible with any proper scrutiny of *φύσις.*

On the other hand, however, we have Parmenides of Elea: a thinker who not only fully embraced the legitimacy of Anaximander's demand for justification, but also understood it to have priority and normative power over the self-affirmation of becoming. The moral demand for justification is in fact taken to an absolutely radical level by Parmenides: he turns it into a *condition for existence.* Whatever there is, it *must* be justified; and whatever cannot be justified should have its being/existence denied.

Parmenides thus seems to let his inquiry be founded upon and primarily guided by the very principles of justification, and not, as in the case of his milesian predecessors and Heraclitus, upon and by the sheer unavoidable self-affirmation of becoming. He dedicates a significant part of his poem to derive the traces that *τό ἐὸν* ("what [there] is") must display from the constraints imposed by what he presumably takes to be "laws", so to speak, in force upon whatever may lay claim to being – laws common to thought and being, for these are, according to him, the same.

The result of such an enterprise is the probably most unusual picture of *φύσις* among all of the so-called *φυσιολόγοι*: a resolute and astonishing denial of becoming. The absurd flux that insisted to happen before his eyes was not conceivable if not as a transgression of the conditions of possibility of being and thought – and thus had to be dismissed as a misleading illusion. In its throne, as the unavoidably true nature of *φύσις,* the eleatic put a static, unchanging, eternal, total, complete, atomic *Being* – the only picture capable of answering to the demands of ultimate justification imposed by thought upon being.

1 Nietzsche, F. Die Philosophie im tragischen Zeitalter der Griechen, §7

The inversion of priority between the demand for justification and the intended object of justification within Parmenides's thought transformed the apologetic attitude of Anaximander into a decidedly and radically challenging one. For after Parmenides, the doors have been opened not only to condemn, but to actually *deny* whatever fails to meet the universal demand for justification bequeathed by Anaximander. And not only did the philosopher of Elea challenge becoming; he has also judged it as having failed to meet the demands of his challenge most unequivocally. The supposed inevitability of the flux was not only solemnly unacknowledged, but substituted by the most absolute conviction of its ultimate impossibility.

Many might claim that acknowledging this fact amounts to the diagnosis that the tragic element of greek thought is strangely absent in Parmenides, but I strongly disagree from that. The inevitability of what is the case is incarnated in his philosophy in a most clear and eloquent way, although it has been radically reinterpreted into a most peculiar expression: that of what tradition would later call *logical necessity*, which abounds in Parmenides's description of reality as in probably no other within pre-socratic philosophy. Existence and the inevitability that adheres to it in tragic ancient Greek thought are thus reshaped as the *result* of the successful and radical satisfaction of the demands of the burden of Anaximander through justification.

Another remarkable consequence of Parmenides little "copernican revolution" in the realm of ancient Greek physiology is that the very nature of the object of inquiry is now also positively determined by the ways and features of the justification given to it. Thereby, justifications become more than mere stamps of approval for an existentially independent object of inquiry, as they were to Anaximander, and are turned into an essential component in the determination of the nature and being of such object.[2]

With Parmenides's radicalisation, then, the contours of the "burden of proof", inaugurated by Anaximander, become significantly less fuzzy and more familiar when scrutinised from a contemporary viewpoint: for now only inasmuch as something can be freed from this burden can it be truly claimed to be (the case). The initial moral questioning is now also existential. So, ever since being (the case) has been deformed into *resisting* the moral

2 a component without which, to say the very least, acknowledging its nature and being is impossible. The importance of this peculiar trait of Parmenides's philosophy to the tradition and issues to be addressed in this work is unparalelled; it is a historical *conditio sine qua non*, a germ of what is called "constructivism" in the philosophy of logic and mathematics, and therefore also for the subject of this dissertation – viz. the question concerning identity of proofs – not to have been doomed never to find a conceptual background against which it actually makes sense.

questioning of Anaximander as radicalised and "perfected" by Parmenides, the originary formal *locus* within philosophical inquiry has been inaugurated which proofs would later occupy – namely, that of final *justification.*

The whole situation depicted so far facilitates a view of Plato as, before anything else, the heir to a clash of titans of thought. On one side, Heraclitus, who accepts and embraces becoming as inaugural with respect to any sense of existence one might come up with, and claims that the truth of *φύσις* – where "truth" is understood as something most faithfully *correspondent* to the nature of *φύσις* – is thus in no sense under the constraints of the laws of thought and being hailed by Parmenides. The sage of Ephesus dances upon the very threshold of madness, demanding of himself and of his listener a disposition to be alphabetised into a language in which enigmas are clear and illuminating and contradictions are sensible and true – the only language he deems rich enough to, in being properly experienced, somehow express, reveal or indicate the truth of *φύσις.* On the other side, Parmenides himself, who claims that what there is and its truth – where "truth" is understood as something ultimately and definitively *justifiable* – have by necessity as their contours, the very outline of their being and nature, the limits imposed by the same laws at which Heraclitus seems to laugh, and thus claims that physiology as a whole, and in fact any investigation whatsoever, is doomed never to attain any truth if not as a result of the strictest observance of such laws. The sage of Elea rests stolid in his fortress of self-discipline, systematically disregarding and almost pitying any expression alien to the only language he deems solid and trustworthy enough to be capable of incarnating and expressing the tragic necessity of being.

The task before Plato seems herculean: nothing less than a reconciliation of Heraclitus's demand for a truth that faithfully depicts and expresses the flux of becoming, and of Parmenides's demand for a truth that is definitively justifiable according to the laws which govern any trustworthy inquiry. Two thousand and five hundred years later, a young Austrian soldier would, while in the trenches and prisons of World War I, come to the conclusion that this task – the reunion of meaningfulness and necessity in truth – is indeed unsurmountable; for it is not, he would claim, herculean, but rather quixotic. But for the entirety of those two and half millennia, Plato's philosophy remained almost unshaken in its role of main paradigm for the proposition of the most various strategies to build a bridge over the abyss between each of these aspects which seem so crucial to the characterisation of truth.

One of the most remarkable pearls of Plato's legacy to the folklore of philosophical tradition – despite what exegetes and historians might have to say about Plato's actual works or intentions – is precisely a model of such a bridge, built as an ontological-epistemic hierarchisation of *φύσις*, epitomised in the famous passage of the Line in the *Republic.* The multiplicity of things which rise and fall within becoming compose the two layers of the domain of *δόξα* (*judgement*, frequently rendered also as *belief* or *opinion*), namely: at the bottom, the images, shadows and *simulacra,* objects of *εἰκασία* (*likeness, depiction* or *figuration,* usually translated as *imagination*); and immediately above, the ordinary physical things of which the images (shadows and *simulacra*) are images (shadows, etc.), objects of *πίστις* (*conviction,* interpreted by e.g. neoplatonists as *faith*, frequently taken to mean *belief* by contemporary readers). Above that of becoming in the hierarchy, Plato posits the realm of the *νοητός* (*intelligible*), domain of *ἐπιστήμη* (especially after Plato, *knowledge*), which is also twofold: at the lower level, the mathematical entities, i.e. relations and proportions between structures, objects of *διάνοια* (*discursive thought,* characterised as, essentially, hypothetic-deductive reasoning); and above it, at the highest ontological level of all, the Forms, objects of *νοῦς* (*thought* or *understanding*, characterised as, roughly, a sort of direct apprehension by the mind, intellectual vision or intuition).

This structure enables Plato to give an account of, first, how stable, immutable being does not rule out the being of the relentless flux of becoming; and secondly, how it is that "trustworthy" truth (as well as the attainment thereof) , dependent on the stable immutability of being as Parmenides has shown it to be, somehow refers to becoming – In other words, it is Plato's means of (a) enabling a post-parmenidean, existential justification of becoming granted that one is ready to "commit parricide" and somehow accept its being; and (b) thus enable a solution to the conflict between Parmenides's and Heraclitus's demands upon truth for, respectively, trustworthiness and adhesion to becoming. By the observed in (a), then, one might feel tempted to suppose that Plato's attitude is to be a return to the apologetic approach to becoming of Anaximander; for indeed, instead of either fully embracing, as Heraclitus, or radically challenging, as Parmenides, the ontological citizenship of becoming, Plato renders the unavoidable self-affirmation of the flux as sufficient to acknowledge that it is not nothing, and therefore does have *some kind* or *degree* of being; and then, by means of his own justification of it, he shows just what kind or degree of being the flux and the things therein have.

Plato's explanation of this hinges essentially upon the notion of "partaking". Just as the images and shadows of *εἰκασία* take their feeble degree of being from the objects of *πίστις*, the latter – that is, the ordinary things which rise and fall within the flux – take their stronger, yet still imperfect and unstable being from the fact that they partake in the stable, perfect and immutable being of the due Forms. Anaximander's morally condemnatory depiction of becoming thus echoes and survives, transformed as it was by Parmenides radical existential appropriation of it, in Plato's characterisation of this ontological realm as existentially imperfect and dependent.

The attainment of truth is then dependent on the "noetic" apprehension of the perfect and independent portion of *φύσις* (which we may, after Plato, call *reality*), i.e. the Forms, which are its source (the term "cause" is more frequently used; meant here is *αἰτία*) and condition of possibility, as well as the source and condition of possibility of the corrupt being of the lower, imperfect layers of reality (cf. e.g. *Phaedo* 100d-e: "*οὐ γὰρ ἔτι τοῦτο διισχυρίζομαι, ἀλλ᾽ ὅτι τῷ καλῷ πάντα τὰ καλὰ γίγνεται καλά."*). This is suggested by Plato as possible by means of the process of dialectic ascension, depicted in the famous passages of the Line (Plat. Rep. 6, 509d – 511e) and the Cave (Plat. Rep. 7, 514a - 517c) in the *Republic*. The attainment of truth in Plato's version of what "trustworthy" inquiry is has thus the property of being, in many respects, *qualitatively informative* with respect to the very *nature and content of φύσις (viz. reality)* as a whole – even, in some sense or degree, with respect to becoming and its inhabitants. Therefore, Plato forges the conception of attainment of truth in inquiry as what tradition would consolidate as *knowledge* – i.e. the grasping of ultimately unwavering, ontologically justified, *necessary* truth that is, simultaneously, *informative,* viz*. non-trivial*. Notice that this not only is a novelty with respect to his predecessors Heraclitus and Parmenides, but also a solution to the aforementioned conflict between them: the first radically rejected and even disdained of the requirement of "trustworthiness" for inquiries towards truth, making of truth and whatever would be equivalent to a "grasping" thereof within his conceptual economy – that is, *if* it makes sense to suppose that there is any place at all for such notions therein – something clearly different from, respectively, something one can know and knowledge; The second, in turn, sacrifices any descriptive, qualitatively informative value of truth with respect to becoming – the reality of which he thus denied, in spite of the plain inevitability of its self-affirmation –, and is contented with having attained a truth that, as trustworthy as it may be, is as empty of content and information as is the being about which it verses, the nature of which is limited to satisfying the necessary and sufficient "formal" conditions to qualify as the only thing that Parmenides's requirements allow to play the role of *φύσις* (the meaning of *φύσις* being thus bent into something close to "what there is"). One of the highlights of Plato's legacy is thus, basically, the creation of the first and, until this very day, paradigmatic conception of knowledge of our tradition and the idea that it is somehow possible.

Now, how does Plato conceive of knowledge? Let us first, as a way to convey the idea of the process of knowledge acquisition outlined by Plato, try to briefly illustrate the aforementioned dialectic ascension, by means of which knowledge is suggested to be achievable. In the immediate approach to the objects within the flux, which happens as *εἰκασία* and *πίστις,* a great many regularities, patterns and proportions may be observed, which cannot however be explained by any further observations made within *εἰκασία* and *πίστις.* In order to explain such regularities, patterns and proportions, an inquirer may assume hypotheses that are capable of leading by necessity to the conclusion that such regularities are true; this step outside of the realm of observable into hypothesising and deducing from hypotheses amounts to the entry into *διάνοια.* At such a stage, capable as the inquirer may already be of recognising true and necessary connections and relations between hypotheses and their consequences, he is still incapable of justifiedly acknowledging the truth of either hypotheses made or conclusions drawn. As Plato explains in the passage of the Line, the soul, by means of dianoetic thought, is forced to investigate "*ἐξ ὑποθέσεων, οὐκ ἐπ᾽ ἀρχὴν πορευομένη ἀλλ᾽ ἐπὶ τελευτήν*" ("from hypotheses, proceeding not towards principles but towards ends") (6, 510b); in short, any explanation attained dianoetically takes unjustified hypotheses as departure point and then heads towards conclusions, which are then true merely relatively to the initial hypotheses. The attainment of truth by means of *διάνοια, πίστις* and *εἰκασία* alone is thus impossible – for, on one hand, unjustified hypotheses cannot justify any conclusions, valid as they may be, drawn from them; and on the other hand, the observation of regularities alone may only go as far as planting *conviction* into the soul of the inquirer, informing nothing regarding *the reason why* such regularities are *true*. To explain how one could overcome *διάνοια* and attain truth by noetic apprehension of Forms – a matter surrounded by mystery and controversy –, one can compare two possible approaches to the hypothetic-deductive reasoning used within *διάνοια* to account for its intended conclusions. Firstly, one could look at it as aiming to directly support/yield the eventual conclusions to which it comes. Secondly, one could look at it as a means to make clear the *reality* from which the hypotheses and the conclusions involved take their eventual truth. The first approach is, again, the one characteristic of *διάνοια*: it merely shows how the eventual truth of the hypotheses and conclusions are deductively related, and how the conclusions can be deduced from the hypotheses, having ultimately nothing to do with whether the hypotheses or conclusions involved are actually true. The second, in turn, aims, to use Plato's expression, "not at ends but at principles"; principles which are nevertheless not to be confused with mere "higher" hypotheses, axioms or something akin. What makes the second approach differ from the first is not a mere inversion of direction in a deductive march between hypothetical assumptions and conclusions, with the ultimate goal of, instead of reaching desired conclusions, finding more fundamental hypotheses that are fit to play the role of principle. Rather, the principles at which the inquirer aims by means of the second approach belong to a totally different category than hypotheses and conclusions – for they are not themselves true, at least not in the sense that hypotheses and conclusions might be true, but rather they are what *makes* true hypotheses and conclusions true: the *realities* to which they correspond, and from which the objects of *εἰκασία* and *πίστις* which eventually motivated them take their defective, imperfect being, by partaking in them. Such realities are the Forms, and their apprehension, *νόησις,* is the necessary and sufficient condition for the actual recognition – rather than the supposition, assumption or conviction – of the truth of any given hypothesis made or conclusion drawn.

Let us consider a concrete example. A theorem – say $((A\supset B)\wedge(B\supset A))\supset\neg((\neg A\wedge B)\vee(A\wedge\neg B))$ – could have its validity explained by someone by appeal to the claim that the due natural deduction rules are valid and the subsequent derivation of the theorem by means of an adequate succession of applications of the rules. Such an explanation can be given in a way that it takes for granted that the inference rules are valid, and merely shows that the theorem can be derived stepwise from them and should be therefore valid. Notwithstanding, the inference rules assumed from which the theorem can be derived could be used in the explanation not as primitive valid principles to be taken for granted, but rather as means to somehow describe, indicate, mirror the *reality* from which they take their validity – say, the fact that conjunction, disjunction, implication and negation are *actually, in reality* such that their meaning is faithfully depicted by the given inference rules. While the first, merely hypothetical use of the inference rules as principles does lead to a valid derivation of a valid theorem from valid principles, it is incapable of accounting satisfactorily for the *reason why* the theorem is valid – it simply shows how it can be properly derived from the unfounded hypotheses that the rules are valid, and the cause (*αἰτία*) of the validity remains thus in oblivion. The second use, in turn, ultimately sustains the conclusion of the validity of the theorem not upon the validity of the inference rules themselves, but rather upon the due portions of reality itself. The reality that makes the rules valid is indicated by means of displaying the rules; in this fashion, each conclusion that is stepwise derived from the initial rules serves as an indication to the stepwise apprehension of the respective regions of reality from where they take their validity, never being inferred in a blind manner, as simply a step the trustworthiness of which is merely inherited from that of the former, but rather always as a sign of the apprehension of the reality that makes it valid. By means of such a process, in which reason makes use only "*εἴδεσιν αὐτοῖς δι᾽ αὐτῶν εἰς αὐτά, καὶ τελευτᾷ εἰς εἴδη*" ("of Forms, [proceeding] through Forms to Forms, and ending in Forms", Plat. Rep. 6.511c), the theorem itself is at last gotten to *know* to be valid.

By means of his solution to the dilemma between Heraclitus and Parmenides, then, Plato has in his own way freed becoming from the moral burden of Anaximander and from the existential threat into which Parmenides has reshaped the latter, and thus established a paradigmatic way of dealing with this cosmological problem. Even more interesting, however, is the fact that, in so doing, he has also provided the means to deal with the ethical-political problem of whether or not justice is possible, both as an ethical and as a political virtue. The possibility of justice is advocated by Socrates in the Republic as secured precisely by the fact that *knowledge* is possible – the cosmological solution, thus, in enabling knowledge, yielded a criterion to what would be a just solution of disputes at the ethical and political level. Indeed, it is not unusual to look at this whole picture upside-down and understand the cosmological solution as somehow tailored to allow for the ethical-political one. In any case, in order to actually and successfully apply viz. *enforce* the ethical and political criterion yielded by Plato's approach, one must be able to *show* that one possesses knowledge, and not e.g. mere appearance thereof – and so, the moral burden of Anaximander is shifted by Plato's philosophy from the shoulders of that which is claimed to *be* (the case) – kosmos and phýsis – to those of that which is claimed to be *knowledge* (judgements, opinions, beliefs, *i.e. δόξα*), thus finally acquiring the epistemic shape of the *burden of proof* as we manipulate it today. Whatever one now claims, now that such a thing as knowledge is possible, one also implicitly claims to *be knowledge*. And this amounts to one's commitment to the capability of justifying not really that which one (at least implicitly) claims to know – i.e. a thing in reality or a fact –, but rather what one claims to be knowledge – i.e. a given judgement, belief, etc. *about* what one knows –, and consequently the very claim to knowledge in the first place. In the present context, such a claim is also a self-entitlement to determining how people should or should not proceed in practical life with respect to the known matter. It is then a matter of the utmost practical importance in this framework whether or not one may have a clear idea of what knowledge itself is – for only against a clear enough idea of knowledge could one indeed differentiate between right and wrong claims to it; thus, between those who have it and those who do not.

Despite the fact that he seems to enable knowledge as a possibility by means of the structure of dialectic ascension as described previously, in no work does Plato present any clear or explicit enough definition of what knowledge is which Socrates embraces. The work in which Plato most directly and substantially discusses this issue is the *Theaetetus.* In this dialogue, after the rejection of the thesis that knowledge is true judgement, the young pupil of Theodorus remembers someone having told him that, when accompanied by an account (“*λόγος”*) – and only then –, true judgement (“*δόξα”*) is knowledge, and proposes this as a new definition of knowledge to be scrutinised:

“ὅ γε ἐγώ, ὦ Σώκρατες, εἰπόντος του ἀκούσας ἐπελελήσμην, νῦν δ᾽ ἐννοῶ: ἔφη δὲ τὴν μὲν μετὰ λόγου ἀληθῆ δόξαν ἐπιστήμην εἶναι, τὴν δὲ ἄλογον ἐκτὸς ἐπιστήμης: καὶ ὧν μὲν μή ἐστι λόγος, οὐκ ἐπιστητὰ εἶναι, οὑτωσὶ καὶ ὀνομάζων, ἃ δ᾽ ἔχει, ἐπιστητά.” (*Theaetetus*, 201c-d)

The passage quoted above is thus the one by means of which Plato, through Theaetetus, eternalises what would become the most influential formula ever used to define knowledge in western philosophical literature: “*μετὰ λόγου ἀληθῆ δόξαν*”.

The proper evaluation of the thesis that the addition of *λόγος* to true judgement yields knowledge depends crucially on, among other things, what one takes *λόγος* to be. It could be argued that Plato himself, in other dialogues, makes some suggestions as to what the nature of a knowledge-producing *λόγος* should be. A conception arguably endorsed by such suggestions is ultimately expressed by the phrase *αἰτίας λογισμός* (cf. *Meno,* 98a), rendered into English as “causal reasoning” or “causal argumentation”; in short, an account of *the reason why the true judgement is true*. If added to the respective true judgement, then, such a *λόγος* would bring about knowledge.

Whether or not Plato actually held such a point of view at any time is an irrelevant matter to our present purposes; in any case, this account of *λόγος*, even if good enough to make a *true* definition of knowledge out of the formula *μετὰ λόγου ἀληθῆ δόξαν,* can be seen as insufficient to the practical purpose of attaining a definition of knowledge that may help in actually sorting out those pieces of *μετὰ λόγου ἀληθῆ δόξαν* which are knowledge and those which are not – and thus also those who possess knowledge about a given matter and those who do not. The reason would be that it is in principle not necessarily the case that there is something in the "surface", so to speak, of a *λόγος* that may provide one with good reason to decide immediately, without further, whether or not it gives an account of the cause of what is claimed to be known. Put another way, one who claims to possess knowledge would always be in a position where it could be asked *how does one know,* if at all, that the *λόγος* added to his true judgement is causal; for whether or not the *λόγος* is an *αἰτίας λογισμός* is in principle not necessarily a self-evident matter. To answer to such questioning, one would in turn have to provide again some *λόγος* to support the truth of the implicitly claimed judgement about the causal nature of the former *λόγος* as a known matter*;* and then the same question could be asked again with respect to the newly provided *λόγος,* and this could go on indefinitely. So, it makes sense to try and spell out a way of reducing the "causal" character of a *λόγος* in this context to some conspicuous, immediately recognisable feature of it, in such a way as to disallow such sort of regress.

The sequel of Plato's Theatetus can of course be variously interpreted, but one interesting way of looking at it is, precisely in the spirit of the observation just made in the previous paragraph, as a tentative investigation of *what form* does a knowledge-producing *λόγος* – arguably, an *αἰτίας λογισμός* – have. We should be aware of the fact that, by employing the terminology "form" now, we are not talking of *the* Forms that Plato posits as the core entities of his ontology; it is rather a loose use of this vocabulary, that should be understood roughly as follows. Inasmuch as the cause of something known – say, that a given body is hot – is given by a *λόγος* in direct dependence on the specific *content* of that which is known – e.g. because of its partaking in hotness, or, less "stupidly", as Socrates suggests in the *Phaedo* (105b-c), because of the fire in it – , this *λόγος* is not necessarily a formal expression of the notion of cause at all; for one has to know something about what hot, hotness and fire are and what relations hold between them to understand why there is even an attempted causal account going on in it, if any. But if there is some way of specifying this *λόγος* as to guarantee that it expresses the cause of something known independently of the content of what is known – say, to say that for everything that can be known, the cause of what is known can always be expressed as an exhaustive enumeration of the fundamental elements of which it is composed –, then there would be a formal constraint upon *λόγος* that would necessarily or sufficiently (ideally both) yield this connection, i.e. necessarily or sufficiently condition its status of *αἰτίας λογισμός*.

Now, the attempted definitions can be roughly described as follows: (a) *λόγος* is to be understood in its plain sense of just speech; (b) *λόγος* is to be understood as an enumeration of the elements of the known matter; (c) *λόγος* is to be understood as exposition of the differential feature of what is known. The precise nature of those attempted definitions need not detain us here; it suffices to notice that all of them are ultimately rejected by Socrates, thus leading the dialogue to end in aporia. Notwithstanding, there are two interesting points to be noticed: first, the formula *μετὰ λόγου ἀληθῆ δόξαν* is itself not positively rejected as a definition of knowledge; and second, the conception of *αἰτίας λογισμός*, to which Socrates seemed quite sympathetic in previous dialogues (cf. *Meno,* 98a) , is not even mentioned as a possible answer to what the nature of a knowledge-producing *λόγος* would be. One could wonder why this is so.

It should be once again stressed that, of course, it might well be the case that *αἰτίας λογισμός* is a clear and maybe even *the* true specification of *λόγος* in *μετὰ λόγου ἀληθῆ δόξαν* according to Plato; but the former observations allow the suggestion that it is nevertheless not exactly to be understood as a *formal* specification thereof – at least not in the same sense as the attempted specifications of the *Theaetetus* are now being claimed to be. The idea is that, even if it were the case that one does have a clear notion of what would be necessary and sufficient conditions for some *μετὰ λόγου ἀληθῆ δόξαν* to *be* knowledge – arguably, that the *λόγος* gives an account of the cause of what is known (e.g. the truth of the judgement, the object of the true judgement, etc.) – , one could still fail to give an account of conspicuous formal restrictions upon *λόγος* which condition its success in accounting for causes – i.e. fail to defend that *λόγος* accounts for the cause by e.g. being an enumeration of the elements or exposition of the differential feature of what is known, whatever it is that is known. So, the *Theaetetus* raises and leaves us with the question of whether or not one could find a proper way of specifying *λόγος formally* so that the formula succeeds in defining knowledge.

If such a picture is adequate, we should see the search of the characters in the *Theaetetus* as directed not merely towards a true definition of knowledge, but rather towards a true definition of knowledge that enables one to *recognise* some *μετὰ λόγου ἀληθῆ δόξαν* as being or not a piece of knowledge by means of scrutinising *conspicuous formal aspects* of the *λόγος* – a much more demanding task, the possibility of completion of which *Theaetetus's* aporetic character puts in doubt. As clear from the former discussion, the political impact of an eventual decisive outcome of such an investigation is of the utmost significance in the framework of Plato's thought – it bears on the possibility of implementation and enforcement of the political model of the *Republic* in practice, and could either take it far beyond or definitely confine it to the status of a mere ideal depiction of political justice.

This view presents Plato as already inaugurating the investigation of the formal aspects of *λόγος* which condition the constitution of knowledge; though in a rather indirect, laconic and unsystematic fashion, rather differently from what Aristotle would later accomplish in the works contained in the *Organon*. The investigation of the nature of *λόγος* in the *Theaetetus* is thus hereby depicted as addressing in essence the same matter as the investigations of Aristotle's *Analytics* – which concern, in the philosopher's own words, *"ἀπόδειξιν καὶ ἐπιστήμης ἀποδεικτικῆς"* (in most translations, *demonstration and demonstrative science* – the translation of *ἐπιστήμη* and related terms in Aristotle, especially in the *Analytics*, is a quite controversial subject; options range from "*science"* through "*knowledge"* to "*understanding"*.) Of course, there must be many solid reasons to reject such a reading of the *Theaetetus* which ignorance and ineptitude prevent me to consider at this stage; but this is again of no importance. The essential point is that Plato's depiction of knowledge-producing *λόγος* as *αἰτίας λογισμός* in dialogues earlier than the *Theaetetus* does invite as relevant possibilities such formal investigations as that of the kind that definitely is carried out in Aristotle's *Analytics* and that may have already started to be developed in the *Theaetetus* – i.e. investigations of conspicuous formal aspects of a *λόγος* that are decisive for it to produce or constitute a piece of knowledge when somehow combined with true judgement.

Be that as it may, the pathway of Aristotle into this terrain had a quite different background than merely Plato's doctrine of Forms, and ended up leading him much deeper into the intricate meanders of what we after his work call formal logic than anyone before him had ever ventured. Let us now, as briefly as possible – which of course causes the sacrifice of fidelity to a significant, yet hopefully still acceptable extent –, try to reconstruct this pathway as it may be seen being trod within some of the investigations of the *Organon.*

It will be useful now to look upon Aristotle's so called logical investigations as devised as a systematic march towards the goal of gathering all the necessary material to explain how *ἀπόδειξιν* and *ἐπιστήμης ἀποδεικτικῆς* may come about by *specifying* these notions – in the quite literal sense of defining by providing gender and specific difference. At the most basic level, thus, we can see three main theses: the so called semantic triangle (as presented in e.g. *De Interpretatione* 16a3-9); the doctrine of the ten categories in its two aspects – the ontological and the logical; and the doctrine of the predicables as presented in the *Topics* (I.v). Abstraction on this last doctrine allows the emergence of the concepts necessary to the definition of, in crescent order of abstractness, modal and plain syllogisms, as presented in the *Prior Analytics.* Finally, provided with the formal tools of syllogism, the philosopher is able to specify the notions of *ἀπόδειξιν* and *ἐπιστήμης ἀποδεικτικῆς* in the *Posterior Analytics,* thus completing his depiction of what formal aspects must a *λόγος* display in order to be knowledge-producing.

The semantic triangle identifies and describes a bond between language, thought and reality. Spoken sounds signify *παθήματα τῆς ψυχῆς* (affections of the soul), as their symbols. Written signs are in turn symbols of spoken sounds – thus being indirectly also symbols of the same affections of the soul. These affections are in turn *ὁμοιώματα* (likenesses) of actual things, which then may, through combination and separation, be true or false – a property which is inherited by their spoken and written signs. The triangle is thus closed by the indirect reference of written and spoken signs to actual things by means of the affections in the soul that they symbolise, which refer to actual things by virtue of their being likenesses thereof.

Now, the doctrine of the ten categories can be looked at as, before anything else, a theory of highest ontological kinds – i.e. substance, quality, quantity, etc. are first and foremost *universals,* real highest genera of things, that thus *are* in the things that belong in them. This is how they are described in the *Categories* (4, 1b25–2a4, Ackrill's translation, my additions in square-brackets):

"Of things said without any combination, each signifies either substance or quantity or qualification or a relative or where or when or (…). To give a rough idea, examples of substance are man, horse; of quantity: four-foot[long], five-foot[long]; of qualification: white, grammatical (...)" [3]

Notice that things said without any combination (arguably the names and verbs of *De Interpretatione* – c.f. 16a9-19) are in this passage not themselves claimed to be either substance or quality or etc. – rather, *what they signify* are. As indicated by the semantic triangle, Aristotle could be referring to two kinds of thing with this: either affections of the soul

3 " Τῶν κατὰ μηδεμίαν συμπλοκὴν λεγομένων ἕκαστον ἤτοι οὐσίαν σημαίνει ἢ ποσὸν ἢ ποιὸν ἢ πρός τι ἢ ποὺ ἢ ποτὲ ἢ (…) Ἔστι δὲ οὐσία μὲν ὡς τύπωι εἰπεῖν οἷον ἄνθρωπος, ἵππος· ποσὸν δὲ οἷον δίπηχυ, τρίπηχυ· ποιὸν δὲ οἷον λευκόν, γραμματικόν· (...)." (Text by Minio-Paluello, L. 1949, *Categoriae et Liber de Interpretatione.* Oxford University Press, London)

or actual things. But the sequel makes it clear that the second option is the adequate one; for the examples given to illustrate substances, etc. – e.g. man, horse, etc. – are referred to as things, and not as likenesses thereof in the soul.

Now, additionally to the highest kinds of beings being such, the semantic triangle gives us a hint that what signifies them – both at the level of the affections of the soul and at the level of spoken and written signs – will somehow mimic their nature. It is thus not at all surprising that in the *Topics,* we find a formulation of the doctrine of the ten categories that applies to the level of predicates/predication (I.9, 103b20–25):

"Next we must define the kinds of categories [predicates, predication] in which the four above-mentioned predicates[predications] are found. They are ten in number: essence ["what-it-is"], quantity, quality, (...). For the accident, the genus, the property and the definition [i.e. the four above-mentioned predicates] will always be in one of these categories; for all propositions made by means of these indicate either essence ["what-it-is"] or quality or quantity or one of the other categories."[4]

Notice that the naming of the first category shifts from *οὐσία* (substance) in the first formulation to *τί ἐστι* ("what-it-is") in the second. This is not by chance: while substance is a category that applies to *beings*, the "what-it-is" applies to predicates; or, what better suits this passage, to the way in which a predicate is predicated of its subject in a *πρότασις* (proposition).

So, just as substance is the primary kind/mode of being, so is "what-it-is" or "(the) this", on the one hand, the corresponding kind of predicate, i.e. the kind of predicate which signifies a substance, and, on the other hand, the kind of relation of a predicate to its subject in a proposition in which the first determines *what* the second is. The same can be said of all categories at the levels of being, predicates and predication. Roughly summarising, then, the doctrine of the categories *qua* formal ontology of the highest kinds/modes of being echoes in the linguistic realm, thanks to Aristotle's semantic/semiotic conceptions expressed in the semantic triangle, *qua* a formal classification of, in the first place, predicates according to the respective beings they signify and, secondly, predications, i.e. ways given predicates that signify given beings are predicated of a subject in a proposition.

4 "Μετὰ τοίνυν ταῦτα δεῖ διορίσασθαι τὰ γένη τῶν κατηγοριῶν, ἐν οἷς ὑπάρχουσιν αἱ ῥηθεῖσαι τέτταρες. ἔστι δὲ ταῦτα τὸν ἀριθμὸν δέκα, τί ἐστι, ποσόν, ποιόν (…) ἀεὶ γὰρ τὸ συμβεβηκὸς καὶ τὸ γένος καὶ τὸ ἴδιον καὶ ὁ ὁρισμὸς ἐν μιᾷ τούτων τῶν κατηγοριῶν ἔσται· πᾶσαι γὰρ αἱ διὰ τούτων προτάσεις ἢ τί ἐστιν ἢ ποιὸν ἢ ποσὸν ἢ τῶν ἄλλων τινὰ κατηγοριῶν σημαίνουσιν." (Text by Bekker, I., Berlin, 1831)

The doctrine of the categories, then, whose logical and ontological aspects are glued together and made indissociable by Aristotle’s semantic/semiotic conceptions, not only bridges ontology and logic but also determines the first as the limiting frame of the Stagirites interest in logical matters, namely, the extent to which they condition declarative discourse (*λόγος ἀποφαντικὸς*) on being.

The basis of Aristotle’s logical considerations is also composed by the doctrine of the four predicables. These are precisely the four predications mentioned in the passage just quoted from the *Topics,* namely: accident (*συμβεβηκός*), *genus* (*γένος*), *proprium* (ἴδιον) and definition (*ὅρος*). As already stated in the very same passage, Aristotle deems these kinds of predication to divide exhaustively into the categories. He explains these four kinds of predication roughly as follows: definitions are predications made by means of an expression (a) which applies to a subject that conversely applies to it; and (b) which signifies the essence (*τὸ τί ἦν εἶναι*) of the subject. *Propria* are those predications which also have property (a), but which do not have property (b), i.e. do not signify the essence of the subject. For example “animal capable of reasoning” is a definition of man, whereas “capable of laughing” is merely a *proprium* of man; for the first not only applies conversely to man, but also signifies its essence, while the second, even though conversely applicable to man, does not signify its essence. *Genera*, in turn, are those predications which do display property (b), but not property (a); instead, while they apply to the subject, the subject does not apply to them. “Animal”, for example, is a *genus* of man. Finally, accidents are those predications which display neither (a) nor (b); i.e., neither can they have the subject predicated of them, nor do they signify the essence of the subject to which they apply. In Aristotle’s own way of expression, an accident is thus something which can belong or not belong to some particular thing. For instance, white, if applicable to wall, is an accident of it; for it could be that any wall was not white – or, alternatively: for neither does wall apply to white (i.e. there are many white things that are not walls), nor does white signify the essence of wall.

As one can easily verify, the four predicables arise from exhausting the possible combinations of a predication possessing or not possessing properties (a) and (b). Since, for any given predication, it either has or has not property (a) – and the same holds for property (b) – , the list of the four predicables is exhaustive with respect to all predications. Sustaining it with a similar reasoning, Aristotle claims this same thesis in the *Topics*. Since, as argued above, the ten categories are also exhaustive with respect to all beings – and also with respect to all predications –, Aristotle then justifiedly claims, in the passage already quoted of the *Topics,* that each of the four predicables belongs to one of the ten categories of predications listed. The bond of soundness and completeness with respect to the formal ontology of highest kinds is therefore not lost in the realm of the predicables.

Abstracting from the four predicables, one reaches the level of modalised predication – i.e. the level of abstraction at which the modal syllogisms of the *Prior Analytics* lie. Predications of *genera, propria* and definitions instantiate, on the one hand, the relation of a predicate *necessarily* belonging to the subject to which it applies, due to possessing property (a) or possessing property (b); Predications of accidents, on the other hand, instantiate the relation of a predicate *two-way possibly* (and hence also *one-way possibly*) belonging to the subject to which it applies, due to possessing neither property (a) nor property (b).

The next higher level – the highest – is thus, unsurprisingly, that which corresponds to the plain syllogisms of the *Prior Analytics;* i.e. the level at which the relation of predication is considered "as such", or at least as such that all of the less abstract classifications of predication described so far instantiate it – and exhaustively so at each respective level of abstraction.

Now, the most interesting and peculiar aspect of Aristotle's treatment of these two last and most abstract levels of consideration of the relation of predication is the fact that it is at them that he starts investigating the formal structure of syllogism (*συλλογισμός*, sometimes alternatively translated simply as "deduction"). The term is not strictly technical throughout the Organon, and it seems to aim at applying to a wide range of valid arguments involving only pieces of *λόγος ἀποφαντικὸς*. Aristotle's definition of syllogism is given at the very beginning of the *Prior Analytics* (A1, 24b18-22) as follows:

"συλλογισμὸς δέ ἐστι λόγος ἐν ὧι τεθέντων τινῶν ἕτερόν τι τῶν κειμένων ἐξ ἀνάγκης συμβαίνει τῶι ταῦτα εἶναι. λέγω δὲ τῶι ταῦτα εἶναι τὸ διὰ ταῦτα συμβαίνειν, τὸ δὲ διὰ ταῦτα συμβαίνειν τὸ μηδενὸς ἔξωθεν ὅρου προσδεῖν πρὸς τὸ γενέσθαι τὸ ἀναγκαῖον." [5]

Despite the fact that such a wide definition is given, Aristotle occupies himself later in the *Analytics* with a very specifically structured group of syllogisms, namely: those having only two premisses and one conclusion, each of which is a (possibly modalised) categorical sentence – in the traditional sense, inspired by the developments of *De Interpretatione –,* with a total of three terms, one of which (the middle) occurs in each premiss but not in the

5 "A syllogism is an argument in which, certain things being posited, something other than what was laid down results by necessity because these things are so. By 'because these things are so' I mean that it results through these, and by 'resulting through these' I mean that no term is required from outside for the necessity to come about." (G. Striker's translation.)

conclusion. This restriction is nevertheless justified by the fact that Aristotle later argues that every syllogism can somehow be given in viz. reduced to this very specific form.

It is therefore in accordance with his definition of syllogism that Aristotle then starts systematically investigating which combinations of two such premisses do and which do not yield syllogisms. In chapters 4 to 6 of book A, for example, the philosopher considers the combinations of plain categorical premisses that yield syllogisms, each of these chapters being dedicated to one of the three figures – the first, in which the middle term is subject in one and predicate in the other premiss; the second, in which the middle term is predicate of both; and the third, in which the middle term is subject of both; thus exhausting all possible combinations of such premisses. Later, the same procedure is carried out with respect to combinations of premisses of which at least one is modalised.

While the interstices of Aristotle's formal investigation of syllogisms are not to be object of our considerations at this moment, it is important not to lose sight of the role of this enterprise in his guiding plan of establishing the features of a *λόγος* which condition its being knowledge-producing. Firmly sustained by the developments at the lower layers of abstraction of his theory of predication – which is in turn itself firmly sustained by his ontology of highest kinds and his ideas on how written and spoken signs of language signify –, Aristotle's investigation of syllogism allows him to concretise his depiction of *ἀπόδειξιν* and *ἐπιστήμης ἀποδεικτικῆς* as his answer to the guiding question in the *Posterior Analytics* (A2, 71b16-25):

"Εἰ μὲν οὖν καὶ ἕτερος ἔστι τοῦ ἐπίστασθαι τρόπος, ὕστερον ἐροῦμεν, φαμὲν δὲ καὶ δι᾽ ἀποδείξεως εἰδέναι. ἀπόδειξιν δὲ λέγω συλλογισμὸν ἐπιστημονικόν· ἐπιστημονικὸν δὲ λέγω καθ᾽ ὃν τῶι ἔχειν αὐτὸν ἐπιστάμεθα. εἰ τοίνυν ἐστὶ τὸ ἐπίστασθαι οἷον ἔθεμεν, ἀνάγκη καὶ τὴν ἀποδεικτικὴν ἐπιστήμην ἐξ ἀληθῶν τ᾽ εἶναι καὶ πρώτων καὶ ἀμέσων καὶ γνωριμωτέρων καὶ προτέρων καὶ αἰτίων τοῦ συμπεράσματος· οὕτω γὰρ ἔσονται καὶ αἱ ἀρχαὶ οἰκεῖαι τοῦ δεικνυμένου. συλλογισμὸς μὲν γὰρ ἔσται καὶ ἄνευ τούτων, ἀπόδειξις δ᾽ οὐκ ἔσται· οὐ γὰρ ποιήσει ἐπιστήμην."[6]

There is indeed at least one kind of "scientific" knowledge to which Aristotle refers in the *Posterior Analytics* that is not demonstrative: knowledge of the first principles. These

6 "Our contention now is that we do at any rate obtain knowledge by demonstration. By demonstration I mean a syllogism which produces scientific knowledge, in other words one which enables us to know by the mere fact that we grasp it. Now if knowledge is such as we have assumed, demonstrative knowledge must proceed from premisses which are true, primary, immediate, better known than, prior to, and causative of the conclusion. On these conditions only will the first principles be properly applicable to the fact which is to be proved. Syllogism indeed will be possible without these conditions, but not demonstration; for the result will not be knowledge." (G. Striker's translation.)

would be, in an analogy, grasped by the soul in a similar way as colours are seen by the eye: just as one is already fully equipped and capable from the outset and needs to learn nothing to see a given colour – say, red – , and yet one cannot get acquainted with the colour without having an adequate experience – e.g. having something red before the eyes; so is one's soul fully capable of immediately grasping viz. learning first principles with the concurrence of proper experiences. Just what these proper experiences would be in the case of first principles is nevertheless not clear.

Given these observations, then, one can see that Aristotle's notion of syllogism of the *Analytics* amounts to the determination of the conspicuous formal aspects – where both "conspicuous" and "formal" have a much stronger, stricter sense than what could be claimed to occur in the *Theaetetus* – of *ἀπόδειξις*, that is, something that can be seen as a fit candidate to the role of *λόγος* in the formula *μετὰ λόγου ἀληθῆ δόξαν* – viz. a knowledge-producing *λόγος.* It seems clear that the idea is rejected that this formula, at least if so understood as discussed here so far viz. as built from a well defined and unequivocal notion *λόγος*, defines knowledge in general; still, obviously enough, it seems adequate to the cases of knowledge that can be called *ἐπιστήμης ἀποδεικτικῆς.*

The notion of a knowledge-producing *λόγος*, then, has been given by Aristotle a special region of interest. While it could be argued whether or not Aristotle's non-demonstrative knowledge demands the concurrence of some sort of *λόγος* in order to come about, not only is it settled that *ἐπιστήμης ἀποδεικτικῆς* demands a specific kind of *λόγος* – namely, demonstration –, but also that the very notion of demonstration does have certain clear and rigorously determined, *conspicuous formal* boundaries incorporated to its "application criteria": they need to be in some sense reducible to the form of syllogism.

This special region of interest determines precisely the extent to which questions concerning the "burden of proof" are made essentially subject to objective, public scrutiny by Aristotle. Notice that *ἀπόδειξις* is not itself completely within the limits of this region; "scientific" proof – or, more generally, what we have been calling knowledge-producing *λόγος* –, thus, and consequently also knowledge in general, escape the strict limits of objective, public surveyability. It seems rather adequate, then, that even while believing that aristocracy is the ideal instance of political justice as a regime, Aristotle bets in democracy as the best political regime one can realistically hope for; the core idea of Plato's *Republic* of enabling justice by grounding it upon knowledge is thus indeed and ultimately relegated to an unenforceable ideal. Looking at the same matter from another angle, though, one could say, equally in consonance with the Stagirite's logical developments, that the essence of Plato's ideal endured and remained through his thought as the conceptual embodiment of justice, and that democracy, inasmuch as it can be said to be part of the legacy of Aristotle's and, more generally, of classical Greek philosophy to ourselves, is nothing but the less imperfect applicable solution to the matter of the form of exertion of political power given certain unavoidable limits imposed by the nature of language, reality and the relation between them.

In opposition to *ἀπόδειξις* and knowledge*,* though, *συλλογισμός* is made intrinsically, unavoidably and very strictly *self-contained* and *conspicuous.* The doors were thus opened by Aristotle to a domain the rules of which are (at the very least in principle) all accessible and clear, but at the cost of it being insufficient to yield or be anyhow more closely related to *knowledge* – and thus, as dictated by the platonic legacy, also incapable of being alone the soil from which any sort of justified ethical or political solution might germinate. Nevertheless, this is far from claiming that this domain is ethically or politically innocuous; indeed, inasmuch as *ἀπόδειξις* is an instance of *συλλογισμός,* knowledge – and hence justice, still following Plato – are at least partially conditioned by the rules discovered by Aristotle in his exploration of the realm of syllogism.

It is also important to notice that the explicit goal of laying the foundations of *ἐπιστήμης ἀποδεικτικῆς* which led Aristotle into this realm in the first place imposed significant restrictions upon his journey therein. Let us not forget that the whole conception of *λόγος ἀποφαντικὸς* which determines the limits of the formal investigations of the *Prior Analytics* is built up from linguistic expressions tailored to mirror ontological entities and structures as to be their names or true or false declarations about them – or, at the very least, selected by virtue of the fact that they do mirror them in such a way. The definition of syllogism of the *Prior Analytics* is nevertheless very broad, and phrased in complete independence of such constraints; the core idea involved in the notion, it seems, is some relation of *consequence* holding of necessity between two distinct and different parts of a *λόγος*– as expressed in the phrase "*a discourse in which, certain things having been supposed, something different from the things supposed results of necessity because these things are so*". Were the goals any different, what instances of syllogism could then have come about? To which extent does it make sense to talk of syllogism in the absence of some or all of such constraints?

The idea of *logical consequence* thus rises within Aristotle's philosophy as an eminently relevant contribution to the founding discussion of which a general picture is sketched here; for it is the core notion of an ontologically/cosmologically based account of how some key formal aspects of language which condition knowledge constrain the settlement of eventual disputes concerning or involving "demonstratively" knowable matters – i.e., in the terms employed previously, it is a depiction of some conspicuous restrictions upon that which may free some claims to knowledge from the burden of proof. But possibly its most fruitful feature is rather the fact that it opened a universe of possibilities less tightly connected to the original, cosmo-ontologically motivated discussion of *φύσις* and more closely related to the posterior, essentially ethical and political debate inaugurated by Plato, namely: how to establish beyond disputability that a given rule is valid? Plato's genius was to try and make of "reality", instead of a mere local criterion for the settlement of disputes concerning its own nature, a general criterion capable of adequately orienting the proper settlement of *every* dispute. But what if *λόγος* itself displays features that, independently of reality and what the eventual relation between them looks like, impose conditions on the proper settlement of at least some disputes? Some crucially important "reactive" streams of thought within philosophical tradition[7] – from traditional medieval nominalism, through Kantian critical philosophy, to Wittgensteinian therapeutic of language – may be seen as arising precisely from the entertainment of some guise of such a possibility; which is certainly invited by the consideration of Aristotle's *Prior Analytics,* especially when carried out in isolation from the rest of his general metaphysical project.

This conceptual field – that of *συλλογισμός,* as determined in terms of the notion of *logical consequence* – is where the main subject of this dissertation is located. We shall here discuss the *identity* of what the literary developments under the labels *general proof theory* and *proof-theoretic semantics* usually call *proofs* – which is really nothing but the same old general notion usually understood as what Aristotle named syllogism, adequately restricted and contextualised. The question concerning the *identity* of proofs is an essential departure, on the one hand, from a discussion which relegates syllogisms/proofs to the role of mere constitutive elements of what may help one get rid of what is, in essence, the same old moral burden by means of which Anaximander inaugurated philosophy – i.e. the role of mere constitutive elements of what may satisfy demands for justification; and, on the other hand, towards the consideration of syllogisms/proofs as having *themselves meaning* and *value*, rather than as mere stamps of approval the only interest of which is to enable certain uses of their conclusions, the nature and meaning of which is fundamentally independent of them. Just what the purpose or relevance of this move is will of course depend on how the meaning

7 One could even reasonably entertain the thought of including Parmenides in such a list.

and value of proofs is to be characterised. Most of the considerations to be made here are thus designed to lead to the questioning of precisely what do different ways of identifying proofs – both considered as a universal kind and as those things which fall under this kind – imply with respect to how one conceives of the meaning and value of proofs; and thus, to the questioning of what do different conceptions of the meaning and value of proofs imply with respect to more general philosophical questions, not only of a logical or epistemic, but also – and most importantly – of a fundamentally *ethical* and *political* character. The eventual failure to appreciate this last point in particular will unavoidably lead to failure to understand the point of this work as a whole.

I. Kant's *Die falsche Spitzfindigkeit der vier syllogistischen Figuren*

About what more precisely are we talking when we ask ourselves about something such as the identity of a proof? Providing an indication as to how it is possible to answer this question, even if at no more than a rudimentary level, in such a way as to make of it the expression of a real and meaningful philosophical issue – and not of some formal or conceptual entanglement pertaining to some specific piece of literature – is the motivating ideal of this section. There is, after all, little to no point in starting to read a philosophical dissertation the nature of the subject-matter of which is to such an extent immerse in mist that one cannot even see why it is at all of any interest – which is the case here. For the present purpose, it might be thus reasonable to start by trying, before anything else, to give an account of the following: *why* is it that one asks (that is, if there is any reason at all to do so) oneself about the identity of proofs?; for knowing the reason why we ask a question might avoid misunderstandings regarding its meaning, its subject-matter, and thus also regarding the adequacy of eventual answers.

It is usual in the literature on identity of proofs – if it makes any sense at all to call something usual in such scarce a literature – either simply not to address questions concerning the reasons why engaging into this sort of investigation, or to quite light-mindedly refer to an *en passant* remark made by Kreisel on the matter in the sixties, which points at how an interesting theory on identity of proofs could arise from the apparently objective terms of discussions among mathematicians interested in priority questions regarding whether or not two proofs are the same. The first attitude is an explicit declination to answering the question; and while the second seems not to avoid it at first glance, it is a mere tergiversation: for it proposes no indication as to why such an investigation could be of profit in any sense. It is also usual to claim that a proper grasp of identity of proofs is of the utmost importance to a better comprehension of the notion of proof itself – and while I cannot bring myself to disagree with such a claim, it does not usually come accompanied by any proper arguments to sustain it, which makes it seem to be put in a quite question-begging fashion.

Instead of trying to provide here arguments to explain why identity of proofs may be relevant to one's comprehension of the notion of proof – which would be an interesting and feasible enterprise –, let us instead begin by taking the shortcut of history, and ostensively pointing at how the matter was actually handled, in the hope of being able to read off of this indications or sensible reasons to approach such a question. The concrete task of this section is thus to advocate that identity of proofs was an issue addressed by mainstream philosophical literature a couple of centuries before general proof theory, proof-theoretic semantics and the akin literary streams that hegemonise the current discussion of the matter even came into being. This shall be done by means of a brief interpretive analysis of Kant's 1762 essay *Die falsche Spitzfindigkeit der vier syllogistischen Figuren*.

One way of formulating the main thesis argued by the philosopher in this opuscule is this: only valid syllogisms of the first figure are to be taken as *"Formeln der deutlichsten Vorstellung eines Vernunftschlusses"*. Before anything else, it is of the utmost importance to make absolutely clear that, whilst holding such a thesis, Kant does not – contrarily to what Patzig quite absurdly claims in his otherwise quite instructive and detailed study of Book A of Aristotle's *Prior Analytics* – by any means question the *validity* of syllogisms in the other three figures. Not, at least, insofar as validity is understood as the *obtention of a relation of (logical) consequence* between premisses at one side and conclusion at the other, and such a relation is in turn understood as the *entailment of the truth of the conclusion by the truth of the premises by force of their mere (logical) form* – a conception which has been the standard on the matter for a considerable time now, and which is quite naturally assumed by Patzig along his text.

Indeed, Kant *shows* in the course of this very text that the validity of syllogisms in the three last figures *follows* (in a quite strong sense, as I hope will be clear in the sequel) from the validity of the syllogisms in the first figure and that of so-called immediate inferences, and explicitly acknowledges such validity when he says that e.g. *"Man kann nicht in Abrede sein, daß in allen diesen vier Figuren richtig geschlossen werden könne."*, and *"Es sind also die übrige drei Schlußarten als Regeln der Vernunftschlüsse überhaupt richtig"* (see §5 of said text). Neither should he be taken to be doing nothing but bursting through open doors, i.e., stating what one could call a sloppy and hardly original version of a logico-conceptual justification for a thesis akin to the Aristotelian distinction between complete (perfect) and incomplete (imperfect) syllogisms – a point which, I hope, will become clear along this section.

Nevertheless, as Patzig states, Kant clearly denies that syllogisms of the three last figures are, as they stand, clearly provided with what he calls "probative force" (*Schlußkraft*) (see ibid.,§4). Of what is he stating that syllogisms belonging to these figures are deprived by means of such a phrase, then, if not validity? Quite luckily, as soon as one tries to answer this question in the obvious way, a great deal is gained. The suggestion I make is thus to take the hypothesis that, even though Kant recognised the validity of syllogisms of second, third and fourth figures, he did not understand them to be *proofs* – that is, he did not believe them to *explicitly display a reason why* their respective conclusions *follow inferentially* or *deductively* from their respective premises. This can be eloquently illustrated by means of a radical example. Take, for instance, an argument which has the axioms of Peano's Arithmetic as premises and Fermat's last theorem as a conclusion: it is very easy to understand that, even though it is valid, it is in no sense whatsoever provided with any probative force – i.e. although it is a valid, truth-preserving inference , it is definitely not a proof. What Kant is claiming about valid syllogisms not belonging to the first figure is that they are analogous to the argument in this example – i.e., they share with it the property of being simultaneously valid and insufficient as proofs. What we should now try to clarify is what Kant understands to be the conditions to be satisfied by an inference so that it enjoys the property of being provided with probative force – for these are the touchstone of the eventual distinction between a valid argument and a proof made by the philosopher, on which this interpretive hypothesis of ours depends.

Roughly speaking, Kant acknowledges two kinds of logically valid inferences: some are inferences of understanding, others are inferences of reason. The first kind corresponds to the inferences that in traditional Aristotelian logic are called "immediate", and that consist of only one premise and one conclusion which follows logically from it. The second kind, also called "mediate" inferences, corresponds roughly to what traditional logic addressed with the theory of syllogism, i.e. two premisses which share a middle term, and one conclusion that logically follows from them which has as its subject a concept that is called "minor term", that also figures in the minor premiss, and as its predicate a concept that is called "major term", that also figures in the major premiss.

Now, the observation of some facts shall help us understand more precisely in what sense Kant understands the conclusions of these kinds of logically valid inference to *follow* from their respective premisses. Thus, regarding the immediate inferences or inferences of understanding, (a) Kant claims that the truth of some judgements is "immediately recognised" (*unmittelbar erkannt*) from some other judgement, "without a middle term". These judgements can be derived from those from which they follow by means of some operation that, in Kant's explanation, does not tamper with the "content" of the premiss – i.e. the concepts involved in the judgement that serves as premiss – but "merely with its form" (see Jäsche *Logik*, §44);

such operations are, in accordance with the form of the premise, simple conversion, conversion by accident, etc. Regarding the inferences of reason, in turn, (b) Kant explicitly states his adhesion to the thesis that so called *dictum de omni* and *dictum de nullo* are *the only* basic principles that justify *all* of them. I quote the opening of §2, where this is stated most clearly:

"die erste und allgemeine Regel aller bejahenden Vernunftschlüsse sei: Ein Merkmal vom Merkmal ist ein Merkmal der Sache selbst (nota notae est etiam nota rei ipsius); von allen verneinenden: Was dem Merkmal eines Dinges widerspricht, widerspricht dem Dinge selbst (repugnans notae repugnat rei ipsi). Keine dieser Regeln ist ferner eines Beweises fähig. Denn ein Beweis ist nur durch einen oder mehr Vernunftschlüsse möglich, die oberste Formel aller Vernunftschlüsse demnach beweisen wollen würde heißen im Zirkel schließen. Allein daß diese Regeln den allgemeinen und letzten Grund aller vernünftigen Schlußart enthalten, erhellet daraus, weil diejenige, die sonst bis daher von allen Logikern vor die erste Regel aller Vernunftschlüsse gehalten worden, den einzigen Grund ihrer Wahrheit aus den unsrigen entlehnen müssen."

Kant then proceeds to show that (b.1) all valid syllogistic moods of the first figure are directly, *immediately* justified exclusively by one of these two principles. Take, for instance, Barbara: all As are Bs, All Bs are Cs; Therefore, all As are Cs. It states that C, a note of all Bs, which is in turn a note of all As, is itself a note of all As. It is thus *par excellence* an example of a single application of *dictum de omni* to a pair of premises that suffices to infer a conclusion. Put another way, one can see *dictum de omni* and *dictum de nullo* as more abstract inference rules, so to speak, which roughly look like the following, respectively: (d.d.o.) Q As are Bs. All Bs are Cs. Therefore R As are Cs; (d.d.n.) Q As are Bs. No Bs are Cs. Therefore R As are not Cs (Where "Q" and "R" are to be understood as variables for "All" or "Some"; Q ≥ R; and finally, All > Some). The first of these rules allows, given a pair of schematic premises of the form All/Some As are Bs, All Bs are Cs – i.e., respectively, the premises of Barbara or Barbari and Darii – to draw the schematic conclusion that All - Some/Some Bs are Cs – i.e., respectively, the conclusions of Barbara, Barbari and Darii. Barbara, Barbari and Darii are thus the schematic rules one "derives"[8] by applying the more abstract schematic rule *dictum de omni,* which is primitively taken as valid. Analogously, one "derives" the rules Celarent,

8 The quotation marks mean only to stress that the operation meant has rules as outputs, not sentences. Such a distinction is in many respects dispensable – see e.g. Schroeder-Heister, P., 1984, *A Natural Extension of Natural Deduction*, The Journal Of Symbolic Logic, Volume 49, Number 4, Dec. –, but I decided to introduce it nonetheless to avoid the reader to take me as having somehow failed to regard it.

Celaront and Ferio by applying *dictum de nullo* to pairs of schematic premises of an adequate form and consequently obtaining a schematic conclusion of an adequate form.[9]

After this, he shows that (b.2) all valid syllogistic moods of the other three figures can be justified by means of these two principles, but only in a *mediate*, indirect way – i.e. the derivation of the conclusion from the premises demands the application of other rules than just *dictum de omni* or *dictum de nullo*. Such additional rules the application of which is demanded by their derivation are those governing the inferences of understanding, i.e., the immediate inferences. The "derivation" of these syllogistic moods consists thus of appropriately connected applications of immediate inferences and d.d.o. or d.d.n., and shall thus inevitably contain more than just three judgements (i.e. two premisses and one conclusion). This means that they belong to a category which Kant calls *ratiocinium hybridum* or *mixed* inferences of reason, as opposed to *ratiocinium purum* or *pure* inferences of reason. Kant's definition of these notions is as follows:

"Wenn nun ein Vernunftschluß nur durch drei Sätze geschieht, nach den Regeln die von jedem Vernunftschlusse nur eben vorgetragen worden, so nenne ich ihn einen reinen Vernunftschluß (ratiocinium purum); ist er aber nur möglich, indem mehr wie drei Urteile mit einander verbunden sind, so ist er ein vermengter Vernunftschluß (ratiocinium hybridum)."

I believe it is thus legitimate to understand as *mixed* the inferences of reason the "derivation" of which by d.d.o. and d.d.n. has more than three judgements; and as *pure*, in turn, those the "derivation" of which from the mentioned principles has only three judgements. This last category, according to Kant, encompasses all and only the valid syllogistic moods of the first figure.

In other words: the valid moods of the first figure are immediately "derived" by single applications of one of two basic principles of the inferences of reason, d.d.o. and d.d.n., to an adequate pair of schematic premises, and the valid moods of the other figures cannot be "derived" in the same way, i.e., by mere single applications of these principles. In a sense, thus, the valid moods of the first figure are "consequences" at the level of schematic inference rules of the mere validity of the *abstract schematic* principles d.d.o. and d.d.n., of which they are direct instances; that is to say, each of them *expresses* solely an aspect of these abstract principles as actual inference rules. Since no valid mood of some other figure instantiates

9 One might find odd that two variables are used in the formulation of d.d.o. and d.d.n., i.e., that the quantity expressions of the minor premise and of the conclusion might be different from each other. This demands the proviso in brackets, that would be otherwise quite unnecessary. This rendering is notwithstanding required for the sake of the coherence of Kant's text, according to which Barbari and Celaront, as valid moods of the first figure, must also be immediately derivable from, respectively, d.d.o. and d.d.n.

directly these abstract schematic principles, they cannot be said to be consequences of the validity of d.d.o and d.d.n. alone – and therefore, they also cannot be said to express them purely.

Here one is already able to notice some striking similarities, both in the subject matters treated and the way they are approached, between the Kantian discussion and that found in general proof-theoretic developments of the early 1970s. Kant is, in effect, first, identifying and separating in a quite formal way what he takes to be the most basic logical operations of deduction; and second, justifying the validity of all logical (chains of) inferences of relevant form by means of establishing their canonisability[10] with respect to such operations. If we consider e.g. Prawitz's 1971 text *Ideas and Results of Proof Theory*, these very tasks are explicitly formulated and their realisation is taken up as part of a kind of general philosophical, foundational program; and despite the adoption of a philosophically engaged attitude along the text and there being a whole section dedicated to historical remarks in it, no reference whatsoever to any work written before 1925 is made. This is remarkable – negatively so –, especially because it seems that one could even argue convincingly that already at the time of Kant, this sort of enterprise was quite old – ancient, indeed – news.

Before we proceed, let us say a few words about the notion of canonisability viz. canonicity, involved in Kant's explanation of the logical validity of syllogisms and also fundamental in paradigmatic approaches of proof-theoretic semantics to what is in essence the same issue. This will give the unfamiliar reader a rough yet useful idea of why this matter has relevance in the present discussion. Let us then suppose that a girl is looking for her book, and she thinks it might be in the drawer of her writing desk. The book, however, is not there, but rather lies upon her bed, hidden under her pillow. There are two ways in which she may properly come to the conclusion that her book is not in the drawer: she may either find this out *directly,* by opening the drawer and seeing that it is indeed not there; or she may find this out *indirectly*, by lifting her pillow and seeing that the book is under it, and thus inferring that it is not in the drawer. *Canonical* verifications are those of the first kind; they are distinguished by the fact that they happen by means of direct reference, so to speak, to that which is established as the conditions to be satisfied so that something – a sentence – may

10 In the sense that an inference is canonisable if and only if it reduces to a canonical one; and an inference is canonical if and only if its "immediate" partial inferences are canonical and its conclusion is obtained by application of one of the immediate inferences or of the first figure inferences of reason – which are taken as primitively valid. We shall still explore further the issue of canonicity in this text – indeed, we shall see that one further, stricter notion of canonicity is mobilised in the text under scrutiny as well, which concerns only inferences of reason.

be justifiedly inferred or simply asserted. Verifications of the second kind, in turn, are only, if at all, indirectly or mediately related to such conditions, and do not – or at least in principle not necessarily – involve reference to them. Now, logically valid inferences or chains of them can be seen as a special case of verifications with respect to their conclusions; thus, in the same sense suggested above, they can be thought of as either canonical/direct or non-canonical/indirect in accordance to how they entail their conclusion, i.e. whether they do so by direct reference to the conditions which make it the case that the conclusion(s) follow(s) from the premiss(es) or otherwise. It is then clear that it makes sense to say that Kant's approach attributes *canonicity* only to valid inferences of understanding and to valid first figure inferences of reason – for the former are "immediately recognised" as valid, and the latter are the only direct instances of the only principles by virtue of which inferences of reason are logically valid, i.e. by virtue of which a conclusion logically follows from a pair of premises in the form of an inference of reason , namely d.d.o. and d.d.n. Furthermore, since it only makes sense to understand the validity of inferences of reason in general as obtaining, if at all, *because* of their reference to d.d.o. and d.d.n., the valid such inferences of the remaining figures must be *shown* to refer to them in an *indirect* fashion to be properly recognised as such, since it is immediately evident that they do not refer directly to these principles. Kant shows this much by exposing the fact that, despite not being canonical, they are *canonisable* – i.e. they can be systematically reformulated into a canonical (chain of) inference(s) of the same conclusion that starts from the same initial premisses –, and therefore indirectly refer to d.d.o. or d.d.n. An inference can be called canonical in this sense if and only if its "immediate" partial inferences are canonical and its conclusion is obtained by (i.e. its last step is an) application of one of the immediate inferences or of the first figure inferences of reason – the "base cases" of canonical inferences for the reasons just explained[11]. Now, while it is well-known that, some two thousand years before, Aristotle had already shown how syllogisms in general can be reduced to first figure syllogisms, it is highly controversial whether or not these reductions are in any sense conceived in the work of the Stagirite as semantical foundations of the validity of the reduced syllogisms, rather than mere didactical devices. In the Kantian text under scrutiny, on the other hand, the first option is unequivocally the case – which shows Kant's anticipation of the strategy of proof-theoretic semanticists such as Prawitz and

11 We shall still explore further the issue of canonicity in this text – indeed, we shall see that one further, stricter notion of canonicity is mobilised in the text under scrutiny as well, which concerns only inferences of reason.

Dummett to explain the logical validity of inferences in general in terms of an epistemically loaded yet logically definable notion of canonicity.

The next capital fact to be noticed regarding Kant's text now is (c) the philosopher entertains the hypothesis that one might think of valid moods of the second, third and fourth figures as expressing aspects of some basic principle of inferences of reason other than the ones expressed by those of the first figure. Not only that, but, furthermore, (d) he believes that the only reason why his predecessors have attributed to the valid moods of the second, third and fourth figures a certain logical "citizenship" is the fact that they did indeed believe them to somehow correspond to such further aspects of basic principles. These two observations are very clearly grounded in the following passage, which, due to its importance, I quote *in extenso*:

"Hier könnte man nun denken, daß darum die drei andere Figuren höchstens unnütze, nicht aber falsch wären. Allein wenn man die Absicht erwägt, in der sie erfunden worden, und noch immer vorgetragen werden, so wird man anders urteilen. Wenn es darauf ankäme, eine Menge von Schlüssen, die unter die Haupturteile gemengt wären, mit diesen so zu verwickeln, daß, indem einige ausgedruckt, andere verschwiegen würden, es viele Kunst kostete, ihre Übereinstimmung mit den Regeln zu schließen zu beurteilen, so würde man wohl eben nicht mehr Figuren,[608] aber doch mehr rätselhafte Schlüsse, die Kopfbrechens genug machen könnten, noch dazu ersinnen können. Es ist aber der Zweck der Logik, nicht zu verwickeln, sondern aufzulösen, nicht verdeckt, sondern augenscheinlich etwas vorzutragen. Daher sollen diese vier Schlußarten einfach, unvermengt, und ohne verdeckte Nebenschlüsse sein, sonst ist ihnen die Freiheit nicht zugestanden, in einem logischen Vortrage als Formeln der deutlichsten Vorstellung eines Vernunftschlusses zu erscheinen. Es ist auch gewiß, daß bis daher alle Logiker sie vor einfache Vernunftschlüsse ohne notwendige Dazwischensetzung von andern Urteilen angesehen haben, sonst würde ihnen niemals dieses Bürgerrecht sein erteilt worden. Es sind also die übrige drei Schlußarten als Regeln der Vernunftschlüsse überhaupt richtig, als solche aber, die einen einfachen und reinen Schluß enthielten, falsch."

Now, bearing in mind that (e) one can easily see that the valid moods of the first figure cover *all* the instances of d.d.o. and d.d.n.; and remembering fact (a), i.e., that Kant understands d.d.o and d.d.n to be the *only* basic principles of inferences of reason; it is then quite easy to understand that (f) the valid moods of the last three figures are not *directly derivable from* or *instances* or *pure expressions* of *any* aspect of *any* basic principle of inference of reason whatsoever; and (g) they are also no expressions, either mediate or immediate, of *any basic principle* of inference *distinct from* d.d.o. and d.d.n. – which can indeed be inferred from (a) alone. Their validity, thus, cannot be explained if not as a consequence of the validity of other schematic inference rules, namely, the immediate ones and those which stem immediately from d.d.o. and d.d.n.; or, put another way, their derivation consists of an adequate composition of the application of basic principles expressed by the valid inferences of understanding and the valid syllogistic moods of the first figure.

We are now, I believe, provided with means to approach satisfactorily the question regarding Kant's notion of "probative force" (*Schlußkraft*) of an argument. One can notice that there is a relation of consequence holding between premises and conclusion in an inference that Kant seems in effect to address primarily in this effort of his, and that is different from the standard, truth-conditionally defined one. That the conclusion *follows* from the premises in this sense, most especially in the case of inferences of reason, means not as much that the truth of the premises entails the truth of the conclusion as, instead, that the conclusion *can be derived from the premises by exclusive means of given licit operations*. Such licit operations are precisely those exclusively by means of which we infer from some judgement or some pair of judgements a further judgement which the competent rules that govern our cognitive apparatus (i.e. *dictum de omni, dictum de nullo*, and the rules governing the inferences of understanding) enable us to *recognise without further* as entailed by the former.

The first important point made by Kant which we should remark here is thus a sort of completeness result concerning this deductive notion of logical consequence and the traditional truth-conditional notion of logical consequence, namely: an inference of reason (or, alternatively, a syllogistic mood) is truth-conditionally valid if and only if it is also deductively valid. So, inasmuch as truth-conditional validity implies deductive validity; and the latter is in turn defined, as noted above, in terms of the *possibility* of deriving a conclusion from given premisses in an *epistemically* binding and absolutely transparent way; then, what is guaranteed by the validity of an argument – understood not only deductively, but also truth-conditionally, since the latter follows from the former – is merely the *possibility* of deriving its conclusion from its premisses in such an epistemically binding and transparent way.

This takes us naturally to a further distinction that must be observed. That something *can* be derived by exclusive means of given epistemically binding operations does not mean that something *was in fact* derived in such a way. Put another way: the fact that an inference is deductively valid does not mean that it displays *in itself* a *sufficient reason why* it is deductively valid. And here lies the condition for an inference to have *Schlußkraft*: it must be *carried out* in such a way that it is *identical* to a complete, self-contained justification of its own validity. Given the way Kant develops these notions in his text, this means that only inferences consisting of adequately connected applications of first figure inferences of reason and inferences of understanding are primitively provided with *Schlußkraft*. All others, even if deductively valid, are all by themselves deprived of *Schlußkraft* – or, in a slightly different formulation, more in accordance with Kant's own explanations, their *Schlußkraft* is indeed the *Schlußkraft* of the respective complete, self-contained justifications of their validity (see ibid. §3, II 52, and §4, II 53,54).

As noted in the beginning of this section, the valid syllogisms of the three last figures *can*, Kant acknowledges, be used for inferring correctly – although they *need* not in any case be used for such a purpose. It is then in a sense completely legitimate to use arguments of these forms – a point that Kant explicitly confirms in the following passage, already transcribed above:

"Man kann nicht in Abrede sein, daß in allen diesen vier Figuren richtig geschlossen werden könne. Nun ist aber unstreitig, daß sie alle, die erste ausgenommen, nur durch einen Umschweif und eingemengte Zwischenschlüsse die Folge bestimmen, und daß eben derselbe Schlußsatz aus dem nämlichen Mittelbegriffe in der ersten Figur rein und unvermengt abfolgen würde. Hier könnte man nun denken, daß darum die drei andere Figuren höchstens unnütze, nicht aber falsch wären."

Nevertheless, insofar as such an argument stands for some epistemically binding and transparent inferential process – i.e. a proof, in the sense which is in accordance with Kant's use of the expression *Schlußkraft* – we must then conclude that it cannot be said to stand for anything *different* from the composition of accordingly epistemically binding and transparent inferential steps – namely inferences of understanding and first figure inferences of reason – that justifies its validity. And the latter are obviously what arguments in the form of first figure syllogisms and immediate inferences stand for. Hence, one must admit that, as *proofs*, valid syllogisms of the three last figures neither are simple, nor can be distinguished from mere compositions of some known simple proofs, namely, immediate inferences and syllogisms of the first figure.

It is thus by means of a strikingly idiosyncratic assessment of the role of the valid syllogistic moods within Logic – that is, that they should first and foremost be the basic and simple formulas for all formal *proofs*, not valid arguments –, which somehow reflects his own idiosyncratic general understanding of the nature of logic and its role for knowledge, as well as his consequent misunderstanding or misjudgement of many of his predecessors works and views on these matters, that Kant is actually giving explicit literary expression – the first to appear that I know of – to an answer to a particular question concerning the *identity of proofs*. Indeed, he is showing that valid syllogisms not belonging to the first figure do not carry with themselves any basic, primitive sort of what Prawitz would 209 years later call "proof idea" that was not already expressed in those of the first figure and in the immediate inferences themselves – thence the "falsche Spitzfindigkeit".

Since only the validity of moods belonging to the first figure *immediately* follows, as Kant claims, from *dictum de omni* and *dictum de nullo* – the validity of all other moods being derivable from these two principles only by means of their reduction to moods of the first figure –, they are alone to be counted as the primitive, simple valid schemes of the inferences of reason. For there is nothing new with respect to basic principles of inferences of reason which is expressed by valid moods of the other three figures; neither some new aspect of *dictum de omni* or *dictum de nullo*, nor some aspect of an eventual further basic principle of inferences of reason – for, according to Kant, there simply is no such principle.

In short, then, Kant is:

(i) not only explicitly justifying the validity, but also establishing the *canonicity* of the valid moods of the first figure based in his acceptance of *dictum de omni* and *dictum de nullo*;

(ii) employing *canonisability* of logical arguments in general as means to justify their validity;

These two tasks – at least the second, at any rate – could be claimed to have already been somehow performed by Aristotle in the *Prior Analytics.* Besides, they do not concern identity of proofs. But further than that, Kant is also:

(iii)demonstrating that valid moods of the last three figures, though canonisable[12], are not canonical at all, by showing that they do not follow immediately from *dictum de omni* or *dictum de nullo*, and by rejecting other eventual basic principles of inferences of reason;

(iv)and thus showing that valid moods of the other three figures, since only valid because ultimately reducible to inferences in the first figure and immediate inferences, do not allow the expression of any canonical argument whatsoever that is *different* from those already expressible by means of valid first figure moods.

I hope to have succeeded in illuminating hereby, then, that identity of proofs, as the contemporary tradition of general proof theory usually mobilises the notion, is a matter addressed by mainstream, traditional philosophical literature since at least 1762; and for reasons that, in being far more philosophically relevant than quarrels among mathematicians worried about questions of priority, do indeed give some indication of the fact that this discussion is intimately connected to how one is to comprehend the notion of proof.

12 In fact, Kant also shows that the fourth syllogistic figure encompasses a valid mood which is not canonisable according to a stricter notion of canonicity, defined only for inferences of reason (see §4, II 55,56). Taking into account what has already been said on the matter of canonicity in this section, one could formulate it as follows: an inference of reason is strictly canonisable if and only if it reduces to a strictly canonical one; and an inference of reason is strictly canonical if and only if all its immediate partial inferences are canonical and its final conclusion is obtained by an application of ddo or ddn. Thus, although all strictly canonical (viz. strictly canonisable) inferences of reason are canonical (viz. canonisable), the converse is not true: the mentioned valid mood of the fourth figure is enough to show this much.

II. Preliminary non-historical considerations

Although Kant's reflections on logic might have actually led him into conceptual meanders where talk of identity of proofs was due, it is not really evident from his approach that a proper answer to the question "what is a proof?" would be jeopardised by an eventual disregard of this matter. We have seen that Kant's interest in identity of proofs had to do with understanding how the semantic and epistemic import of valid syllogistic moods of distinct figures compares to one another. His conclusion is epitomised by the title of his short investigation: the distinction between four syllogistic figures is understood by him as a "false subtlety", in the sense that the valid moods belonging to the three last figures do not bring about probative import of a different semantic or epistemic nature from those belonging to the first. In a sentence: valid "inferences of reason" in the last three figures are, *as proofs,* identical with compositions of apropriate "inferences of understanding" and inferences of reason in the first figure. But how could such a development be in any way regarded as contributing to Kant's better understanding of what a proof is? Indeed, it seems that we witness the exact reverse happening in this effort of Kant: it looks as if Kant's rather clear previous ideas regarding of what proofs are – crucially based in the conviction that *dictum de omni* and *dictum de nullo* are the only principles which justify the validity of inferences of reason – were indeed what provided him with the material to understand why they behave the way they do with respect to their identity. So, as much as this short historical consideration might have exemplified and illustrated the fact that the philosophical investigation of identity of proofs may be relevantly influenced by one's eventual way of previously conceiving the notion of proof, then, it certainly did not give an account of why it would be an indispensable or at least a particularly important enterprise to the proper investigation and understanding of what a proof is; neither in Kant, nor in other cases, let alone in general. The question then returns, now begging for a different approach: why is it that one asks (that is, if there is any reason at all to do so) oneself about the identity of proofs? As probably clear by now, however, this question itself admits various interpretations, to which correspond accordingly different answers – which are in turn justified by reference to different phenomena.

Firstly, it should be observed that one can frequently prove or argue validly for the same things in a variety of significantly different ways. It is thus convenient to put the matter in the following rough yet expressive enough terms: it seems clear that the identity of a proof is not entirely determined by *what* it establishes – e.g. some consequence relation between propositions or sets of them –, but also by *how* it does this much. This, however, by no means implies that every two different ways of proving something are *significantly* different from one another – which means that it remains rather unclear which differences between two given proofs of a certain thing are significant with respect to their identity and which are not.

Secondly, one should also notice that there is no reason to rule out without further the possibility of proofs of different things being in such a way analogous that their differences are not to be regarded as significant. Therefore, it also remains unclear whether two proofs must be deemed significantly different by force of the mere fact that they prove different things.

Bearing these two observations in mind, the question as to why one asks about the identity of proofs can be answered in at least two versions. A brief research on the specialised literature (see e.g. Widebäck 2001, p.9) shows that one *does* ask oneself about identity of proofs because of what the first observation points at, namely, that what a proof proves – which is something mostly taken for unequivocal, transparent and undisputed – is not evidently *enough* to determine its identity. Even if tacitly, though, it is more often than not assumed that two proofs are significantly different whenever they are proofs of different things, and more precisely *because* of the mere fact that they are proofs of different things. The second observation points out that this move lacks justification, and therefore gives us a further reason why one *could* ask oneself about the identity of proofs – even though it is not a reason that has taken many to actually do it –, namely, because it might well be the case that what a proof proves is not even a necessary trait to the determination of its identity. In other words: what one usually takes for granted and plainly transparent about given proofs , namely *what* they are proofs of, is neither sufficient nor (necessarily) necessary to assess whether the proofs considered are significantly different or not; and these are reasons why one respectively does and could ask oneself about the identity of proofs.

Still, these answers are of little use in the task of understanding what importance identity of proofs has in the investigation of the notion of proof, since they only help us understand that the identity of a proof, whatever this means, is not something easily, clearly or even necessarily understandable in terms of things that are purportedly already clear, such as e.g. what a proof is a proof of. Yet there is still a version of the persistent question of the first paragraph – the answer to which could in fact be more useful to understand why identity of proofs is relevant to the understanding of what a proof is – that remains unanswered, namely: why *should* one ask oneself about the identity of proofs?

The first points to be fixed in order to avoid getting lost in dealing with this question here are then these: firstly, since we want to understand how identity of proofs may contribute to our understanding of what a proof is – if at all – , we should not have to rely on a previously fixed conception of what a proof is in order to adequately account for the question of identity of proofs; and secondly, we shall need to formulate more clearly just what is being investigated here under the label *identity of proofs*. It is no sheer stipulation, but rather the short study of Kant's text on the syllogistic inferences just carried out, toghether with the observations made in the previous paragraphs that provide us with a way of defining this: under an investigation concerning identity of proofs, we understand a study of the conditions under which the semantic or epistemic values of things which are understood to be proofs are the same viz. equivalent.

There is, of course, a great many ways of answering the question as to why one should investigate identity of proofs , which shall usually vary in accordance with ideological inclinations and goals. I here merely suggest a sketch of a possible answer that, I hope, shall suffice for now. A brief explanation of this may be given by appeal to the general Quinean slogan "no entity without identity". In a brief text on the concept of identity[13], Sundholm points out the following:

" Consider the two types $\mathbb{N}^{+} \times \mathbb{N}^{+}$ of ordered pairs of positive integers and $\mathbb{Q}^{+}$ of positive rationals. Formally they have the same application criterion:

$$\frac{p : \mathbb{N}^{+} \quad q : \mathbb{N}^{+}}{\langle p, q \rangle : \alpha}$$

<2,3> and <4, 6> are equal elements of type $\mathbb{Q}^{+}$, but not of the type $\mathbb{N}^{+} \times \mathbb{N}^{+}$. In order to individuate the types in question different criteria of identity are needed: the type $\mathbb{N}^{+} \times \mathbb{N}^{+}$ is individuated by the identity criterion

$$\frac{p : \mathbb{N}^{+} \quad q : \mathbb{N}^{+} \quad r : \mathbb{N}^{+} \quad s : \mathbb{N}^{+} \qquad p = r : \mathbb{N}^{+} \quad q = s : \mathbb{N}^{+}}{\langle p, q \rangle = \langle r, s \rangle : \mathbb{N}^{+}}$$

and the type $\mathbb{Q}^{+}$ by the criterion

13 Sundholm, G.B. (1999) Identity: Propositional, criterial, Absolute, in The Logica 1998 Yearbook, Filosofia Publishers, Czech Academy of Science, Prague, pp. 20–26.

$$\frac{p:\mathbb{N}^{+} \quad q:\mathbb{N}^{+} \quad r:\mathbb{N}^{+} \quad s:\mathbb{N}^{+} \qquad p\times s = q\times r:\mathbb{N}^{+}}{\langle p,q\rangle=\langle r,s\rangle:\mathbb{Q}^{+}} \text{ „}$$

The same observation could be applied to the present context of discussion of proofs: we have an application criterion (say, being [represented by/expressed as/carried out by means of] a derivation), out of which a distinct, particular notion of proof (as a type) could only be made in case a specific identity criterion is associated with it, different identity criteria yielding correspondent altogether different notions of proof. Notice that an identity criterion is necessary not only to determine the identity of the individual proofs inside the type, but also the identity of the very type itself. It is in this sense that Martin-Löf, according to Sundholm, suggests as an easy way to reconstruct the Quinean slogan “no entity without identity” the combination of the type-theoretic maxims “no entity without type” and “no type without identity”. One could understand the claims that identity of proofs is a central question to the field of general proof theory and, more generally, to the task of providing an answer to the question “what is a proof?” rather in such a spirit.

1. Proofs, results and identity criteria

It has already been indicated here how it is that accounting for the identity of proofs may be regarded as a fundamental task for the clarification of the notion of proof. But there is something we may say about proofs – at least as far as they concern us in this work – already at this initial point, without being afraid of betraying our guideline and determining what a proof is beforehand: proofs are always proofs *of something.* This transitive structure of proofs makes it clear that every proof inevitably brings with itself also *that which is proved*: which we shall henceforth call the *result* of the proof.

Now, depending on how one chooses to deal with certain matters which are relevant for the determination of the semantical value[14] of a proof, accordingly different criteria of identity of proofs may be deemed adequate viz. inadequate. Here, three such matters shall concern us: first and most importantly, (a) the relationship between the identity of a proof and the identity of its result; secondly (b) the category to which proofs are taken to belong; and

14 This expression should be taken in a general sense here, which may encompass things such as meaning, sense, intension, extension etc.; and not simply as some denotation.

finally (c) the category to which results of proofs are taken to belong. Variations in how one conceives of (a) will generate distinct general kinds of criteria of identity of proofs; in the case of (b) and (c), in turn, variations will affect the meaning of certain formal consequences of adopting certain given criteria of identity of proofs.

In the sequel, a taxonomy will be proposed for each of the items (a) – (c). Provided with this tool, we will be able to identify and better understand certain conceptual entanglements concerning the discussion of proposals as to how one could account for equivalence between derivations and how they fare as attempts to describe or deal with identity of proofs.

a. Taxonomy of identity criteria

In some possible frameworks, identity of proofs shows itself as a rather trivial question, unworthy of our attention. For instance, one could simply regard all proofs as equivalent inasmuch as one considers relevant merely an aspect that they all trivially share, namely: being a proof. Since their semantic or epistemic role is analogous if seen from this perspective, there is no sense in separating proofs by means of such a criterion. Unsatisfactory as it may be in a number of senses, this is still a possible and quite sufficient way of accounting for the identity of proofs: to deem the relation as perfectly trivial.

A less trivial way of dealing with the matter is the folkloric background conception the rejection of which seems to have triggered the young literature dedicated to identity of proofs, namely: what is relevant about proofs is ultimately *what* they prove, two proofs being thus equivalent if and only if they prove the same thing. Of course, the question of just what it is that a proof proves viz. what is it that we understand to be the *result* of a proof is to be answered satisfactorily if any clarity is to come from such an attitude towards identity of proofs – a question mostly neglected by discussions within this young literature, by the way. In any case, this attitude towards identity of proofs is tantamount to reducing the question regarding the identity of a proof to one regarding the identity of whatever it is that we consider as the result of a proof; in other words, it is a trivialisation of the identity of a proof with respect to that of its result. This is, then, a second possible way of accounting quite sufficiently for the identity of proofs while dismissing the question itself as ultimately uninteresting: to reduce the identity of a proof to that of its result, thus resolving the initial question in terms of one that does not necessarily have anything whatsoever to do with proofs.

This has nevertheless proved to be a quite unsatisfactory thesis for the enthusiasts of so-called general proof theory, as I have just hinted at. An emblematic expression of this fact is given by Kosta Došen 2003 (p.14):

"For the whole field of general proof theory to make sense, and in particular for considering the question of identity criteria for proofs, we should not have that any two derivations with the same assumptions and conclusion are equivalent, i.e. it should not be the case that there is never more than one proof with given assumptions and a given conclusion. Otherwise, our field would be trivial."

We shall return to this observation of Došen and discuss an interesting aspect of it soon; for now, it suffices to note that it commits to the idea that there being different proofs of at least some given result is a necessary condition for an adequate approach to the notion of proof.

Now, this implies a rupture with one of the directions of the precedent thesis, but not with the other – i.e. proving the same result is not anymore considered a sufficient condition for two proofs to be equivalent, but it may well remain as a necessary condition for this much. And it seems to be precisely the decision to regard having the same result as a not sufficient yet necessary condition for two proofs to be equivalent that motivates most of the actual developments on identity of proofs available. This is the first attitude towards identity of proofs considered here so far that does not trivialise or dismiss the question in any sense, viz. the identity value of a proof is made neither trivial nor reducible to that of something else. Nevertheless, the identity of the result still plays a prominent role in the determination of the identity of the proof in this view, and an account of it must be provided so that the limitation it imposes upon the identity of the proof becomes clear.

One could still of course move a step further and conceive the possibility of proofs of different results being equivalent, thus breaking also with the other direction of the thesis criticised by general proof theory. This indeed hardly could be regarded as an extravagant hypothesis; for proofs of different things might be strongly analogous in various and significant senses. In such a framework, it is neither necessary nor sufficient that two proofs share the same result for them to be equivalent – i.e. there may be distinct proofs of one and the same result, and there may also be equal proofs of distinct results. At least in the outset of this conception, then, the identity of a proof is neither trivialised, nor reduced to the identity of something else, nor restricted in any decisive way by the identity of its result.

These different possible ways of approaching identity of proofs – and, consequently, proofs – allow us to speak of a semantical taxonomy of criteria for identity of proofs. The exposition I shall now provide of such a taxonomy shall prove helpful, and I hope to be able to make that clear, in avoiding confusion with respect to the proper philosophical evaluation of different formal proposals which intend to deal with the matter of identity of proofs, as well as how they are related to each other conceptually.

With respect to the relation between a proof and the result of a proof, criteria for the identity of proofs let themselves divide exhaustively in the following categories:

1) Trivial criterion restricted to the result: the second kind of criterion described in this section, which identifies all and only proofs of a same result, i.e. which reduces the identity of proofs to that of their result;

2) Non-trivial criterion restricted to the result: the third kind of criterion described in this section, which identifies only but not all proofs of a same result. To this kind of criterion belongs e.g. what we may call the strict criterion: a kind of criterion not described above, which simply deems every proof equivalent only to itself;

3) The third criterion: a kind of criterion not explicitly described above, which identifies all but not only proofs of a same result. To this kind of criterion belongs e.g. the first criterion described in this session – which we may call the unrestricted trivial criterion –, which simply deems all proofs equivalent, regardless of their result;

4) Unrestricted non-trivial criterion: the last kind of criterion described above, which identifies neither all nor only proofs of a same result;

The importance of stressing and not losing sight of the fact that this taxonomy is of a *semantical* nature could never be overestimated, so this is to be stated now, before any confusion on this matter finds opportunity to come about: within this section, until this point, the expressions “proof” and “result” all have an essentially informal, semantical meaning; so “proof” does not mean derivation, “result” does not mean end-formula, etc. This sort of assimilation of meaning which we are blocking here is, by the way, precisely what seems to happen at the transcribed passage of Došen 2003 above: assumptions and conclusions of derivations are *formulas*, just as syntactical as these are; assumptions and conclusions of *proofs*, on the other hand, are *semantical*. Thus, the triviality viz. non-triviality of a relation of equivalence between *derivations* of a certain A from a certain Γ is neither a necessary nor a sufficient condition for the triviality viz. non-triviality of any relation of equivalence between *proofs* of a certain conclusion from given assumptions – unless, of course, one shows that there is a correspondence of a specific nature between, on the syntactical side, formulas, and, on the semantical, results of proofs. In principle, it could well be the case that, for every A and Γ, all derivations of A from Γ were equivalent to one another, and yet there still were different proofs of a given result from given assumptions – just let the distinct formulas A and B express the same proof result and the distinct sets of formulas Γ and Δ express the same proof assumptions, and further let no derivation of A from Γ be equivalent to a derivation of B from Δ, and *voilà*. Since we have argued neither for nor against any kind of correspondence relation between formulas and proof results, and since it is also fairly usual to see some such correspondence being taken for granted in the literature, the proviso just made is justified.

b. Taxonomy of the notion of result of a proof

Another aspect of our taxonomy shall now concern how the notion of *result of a proof* is understood in the framework of a given criterion for the identity of proofs. As mentioned above, this and other crucial questions are left mostly in the oblivion within the discussion of identity of proofs.

Given that proofs are being dealt with here only inasmuch as they can be carried out by means of derivations, it is not to take too narrow a perspective if we deal with whatever it is that we take to be results of proofs as belonging either to a propositional or to a deductive category.

1) Propositional category
1.1 Result of a proof as a (collection of) proposition(s);
1.2 Result of a proof as a(n) (collection of) intension(s) stricter than a proposition – say, e.g., (scheme of) formulation of a proposition;
1.3 Result of a proof as a(n) (collection of) intension(s) looser than a proposition – say, e.g., form/scheme of proposition

2) Deductive category

2.1 Result of a proof as a (collection of) deduction(s);

2.2 Result of a proof as a(n) (collection of) intension(s) stricter than a deduction – say, e.g., (scheme of) formulation of a deduction;

2.3 Result of a proof as a(n) (collection of) intension(s) looser than a deduction – say, e.g., form/scheme of deduction.

c. Taxonomy of the notion of proof

Proofs themselves, in turn, given the restrictions imposed by the application criteria mentioned just above, can be here regarded as belonging to one of the following categories:

1) Proofs as *performable* ; i.e., e.g. as a more or less specific list of instructions that can be carried out;

2) Proof as *performance*; i.e. as the act of carrying out such list of instructions;

3) Proof as *performed*; i.e. e.g. as the object produced by the such a performance viz. by carrying out such instructions.

It is never too much to stress again that the taxonomic divisions of this subsection are by no means intended to be exhaustive with respect to any notion of proof broader than one that can be outlined by means of the application criteria mentioned above, i.e. one such that all proofs are expressible as derivations. Even when considered in this narrower framework, it is certainly nothing but one among many possibilities of categorising proofs in a sufficiently systematic way – although one the conceptual nature of which will allow for the proper illumination of certain aspects of proofs, as well as of attempts to account for their identity, ahead in this work.

Unlike the notions addressed in the previous two subsections of this taxonomy, the notion of proof has been addressed frequently enough by taxonomic approaches similar to the one presented here (see e.g. Sundholm 2000 and references mentioned therein). In the absence of a simultaneous taxonomic account of the other notions addressed here, however, this sort of development would not be enough for a proper evaluation of relevant aspects of the question concerning the identity of proofs, as will become clear along what is to follow.

III. A philosophical assessment of the normalisation thesis concerning identity of proofs

The kind of task to be carried out in this section can be seen as belonging to, quoting Kreisel, *"the sort of* Kleinarbeit *which is generally needed to support a genuine hypothesis (...) as opposed to a mere mathematical fancy."* (*A Survey of Proof Theory II*, p.114, in *Proceedings of the Second Scandinavian Logic Symposium,* Edited by J.E. Fenstad, Studies in Logic and the Foundations of Mathematics Volume 63, Elsevier ,1971, Pages 109-170). For reasons which will, I hope, become clear in the course of the text, substituting "support" for "understand" and "genuine hypothesis" for "relevant proposal" in Kreisel's formulation would make it more in tune with the perspective of this work. We shall examine the *normalisation thesis* concerning identity of proofs with respect to some of its significant philosophical aspects: what does it state?; to which question does it propose an answer?; why, if at all, is this question relevant?; how good an answer to such a question is it?; how, if at all, can it influence in the handling of other relevant problems?; etc. Thereby, we shall try and evaluate, against the background of the reflections already carried out, how this influential proposal on identity of proofs fares in the task which is allegedly the very goal of so-called general proof theory: to improve our understanding of what a proof is. The sequel of this text requires that the reader is acquainted with some basic notions of the framework of natural deduction formalism – most especially that of *derivation* (or, alternatively, deduction) and those directly related to it. The unfamiliar reader is thus referred to the first chapter of Prawitz 1965 doctoral dissertation *Natural Deduction,* most especially §§ 2-3.

1. Some essential traits of the normalisation thesis

The normalisation thesis "official" formulation is the one given to it by Prawitz in his 1971 *Ideas and Results in Proof Theory,* p.257 : "Two derivations represent the same proof if and only if they are equivalent"[15]. As dull and tautological as it may sound put this way, this thesis is, in several respects, not trivial at all.

15 Prawitz, D. (1971) *Ideas and Results in Proof Theory* in: J.E. Fenstad ed., Proceedings of the Second Scandinavian Logic Symposium, North-Holland, Amsterdam (1971), pp. 235-307.

Firstly, the notion of equivalence in terms of which it is formulated is a very specific one: it concerns natural deduction derivations, and can be defined as the reflexive, transitive and symmetric closure of the relation of immediate reducibility between derivations. The reductions considered are those involved in the normalisation of derivations; thus, a derivation reduces immediately to another derivation (see *Ideas and Results,* Section II.3.3) when the latter is obtained from the former by removing a maximum formula (i.e. a formula with a connective * that is the conclusion of an introduction of * and the major premiss of an elimination of *). For example, in the case of conjunction, the derivation:

$$\dfrac{\dfrac{\begin{matrix}\Pi_1 \\ A\end{matrix} \quad \begin{matrix}\Pi_2 \\ B\end{matrix}}{A \wedge B}}{\begin{matrix}A \\ \Pi_3\end{matrix}}$$

Immediately reduces to the derivation:

$$\begin{matrix}\Pi_1 \\ A \\ \Pi_3\end{matrix}$$

Derivations also reduce immediately to others by immediate expansions. These are reductions that can be performed on the minimum formulas of normal derivations, thus conforming them to what Prawitz calls expanded normal form, where all the minimum formulas are atomic. For example, let the following derivation be normal and $A \wedge B$ be a minimum formula in it:

$$\begin{matrix}\Pi_1 \\ A \wedge B \\ \Pi_2\end{matrix}$$

By immediate expansion, it reduces immediately to the derivation below[16]:

$$\dfrac{\dfrac{\begin{matrix}\Pi_1 \\ A \wedge B\end{matrix}}{A} \quad \dfrac{\begin{matrix}\Pi_1 \\ A \wedge B\end{matrix}}{B}}{\begin{matrix}A \wedge B \\ \Pi_2\end{matrix}}$$

16 One might ask oneself why the application of immediate expansions is restricted to these cases. The reason, as I hope will become clear, has nothing whatsoever to do with identity of proofs, but rather with *normalisation* of derivations; which is the very purpose for which these reductions were devised in the first place. It turns out that the restriction upon immediate expansions is what allows the obtention of strong normalisation in the presence of the reductions by means of which maximum formulas are removed (on this and closely related matters, see C.Jay, N.Ghani, *The virtues of eta-expansion.* J. Functional Programming 5 (2): 135-154, April 1995, Cambridge University Press).

Let us note in passing that there is a certain analogy between the two kinds of reduction just presented, namely: both involve derivations in which one draws, from a given premiss, a conclusion which plays no role, so to speak, in the obtention of the end-formula from the top-ones – in some cases, such as that of conjunction, one even immediately *returns* to the said premiss after drawing the kind of conclusion in question. Thus, in the first case, the derivation from Γ to *A* to be reduced is characterised by the introduction and immediately subsequent elimination of a complex formula, which is *removed* by means of the reduction without prejudice to the derivation of *A* from Γ; and dually, in the second case, the reduced derivation from Γ to *A* is characterised by the elimination and immediately subsequent introduction of a complex formula, which comes about by the *insertion* of one or more simpler formulas by means of the reduction.[17]

Other reductions by means of which a derivation immediately reduces to another are the so-called permutative reductions, which concern the eliminations of disjunction and of the existential quantifier. By means of them, it is possible to remove maximum segments (which are called "maximum" by a similar reason as maximum formulas are: they are sequences of repeated occurrences of a same given formula in a row viz. immediately below each other, such that the first one is the conclusion of an application of an introduction rule and the last one is the major premiss of an application of an elimination rule. Notice that maximum formula and maximum segment can be so defined that the first are a special, limiting case of the latter, where only one occurrence of the given formula happens. Longer maximum segments may come about by virtue of applications of the elimination rules of disjunction and existential quantifier. See Prawitz 1965 p. 49, and Prawitz 1971, II.3, p.248 3.1.2) . There are also the reductions called immediate simplifications, which aim at removing eliminations of disjunction where no hypothesis is discharged; there are further similar immediate simplifications that concern the existential quantifier, and also so-called "redundant" applications of the classical absurdity rule.

Thanks to the analogy discovered between natural deduction and typed lambda-calculus known as the Curry-Howard correspondence – especially effective in the

17 Here one might also wonder why the second reduction, contrarily to the first, inserts, instead of removing, a formula which is dispensable for the obtention of the end-formula of a derivation from its top-ones. The answer, once again, is not connected to identity of proofs, but rather to the *confluence* of the reduction procedure: if we took the immediate expansions in the other direction – i.e. as contractions rather than expansions –, then e.g. the unicity of the normal form of derivations in general would be lost in the presence of the reductions by means of which one removes maximum formulas. Again, see C.Jay, N.Ghani, *The virtues of eta-expansion.* J. Functional Programming 5 (2): 135-154, April 1995, Cambridge University Press).

conjunction-implication fragment –, it is possible to verify that the equivalence relation yielded by the reduction system described here and in terms of which the normalisation thesis is formulated corresponds to βη-equivalence, and to β-equivalence in case immediate reductions are left out of the system. The normalisation thesis can thus be formally regarded as the identity clause of a definition of "proof" as a type, the central idea of which is that β- and η-conversions – or, correspondently, that the conversions respectively associated to the reductions that eliminate maximum formulas and to the immediate expansions – are identity preserving.[18]

It should also be observed that the thesis puts forward two separate claims: one to the *soundness,* another to the *completeness* of the equivalence relation defined in terms of the reductions with respect to the informal relation holding among derivations representing same proof. The *soundness* or *if* part says that equivalence *suffices* for two derivations to represent the same proof, i.e. that all derivations equivalent to one another represent the same proof, or, in other words, that the formal relation of equivalence between derivations is *sound* with respect to the informal relation holding among derivations representing same proof. The *completeness* or *only if* part of the thesis, in turn, says that equivalence is *necessary* for two derivations to represent the same proof, i.e. that there are no derivations non-equivalent to one another that represent the same proof, or, in other words, that the formal relation of equivalence between derivations is *complete* with respect to the informal relation holding among derivations representing same proof.

Rather than an arbitrarily conceived criterion, this idea is inspired to a significant extent by simple and in some senses appealing philosophical conceptions and formal results, as we shall try to make clear in what follows.

a. Informal idea and relevant (arguably supporting) formal results

In accordance with the initial assumptions made regarding how one manipulates the notions of *proof*, *derivation* and the relation between these in a semantical framework, the normalisation thesis seems to make sense as an answer to the question regarding identity of

18 For brevity, the terminology of the lambda-calculus will be frequently employed when referring to the reductions and conversions, even though the discussion here takes place in a natural deduction setting. This is especially unproblematic here since, as will be clear, the conjunction-implication fragment is enough for almost all considerations to be made.

proofs viz. synonymy of derivations if understood as a formal account of a very simple and reasonable idea:

(α) that any two proofs the difference between which resolves into *irrelevant features* are indeed not significantly different;

(β) that any two proofs that differ with respect to any other feature are significantly different.

To this, advocates of the thesis usually add that

(γ) a specific proof can always be given with no *irrelevant features.*

In short, the idea is that any two proofs are significantly different if and only if they are different *even when given without any irrelevant features*. Back reference to these three clauses by means of the respective tags (α), (β) and (γ) will be abundant in the sequel.

Besides, the following three formal results seem crucial, both from the historical and the conceptual viewpoint, to the proposal now under scrutiny, namely: the normal form theorem, the (strong) normalisation theorem and the uniqueness of normal form.

The *normal form* theorem states that every formula A that can be derived from a set of formulas Γ can be derived from Γ normally (that is, without the occurrence of maximum formulas viz. segments in the derivation); i.e., every valid consequence relation between some Γ and some A is provable without resort to any non-normal derivation. This result can be regarded as a kind of completeness theorem that seems essential to the normalisation thesis, which could be stated in the following fashion: every provable *result* can be proved normally; or: normal derivations alone can prove all provable *results*. Furthermore, the *normalisation theorem* yields a mechanical procedure by means of which derivations can be reduced to normal ones, showing that the (syntactical) difference between them resolves into certain deductive patterns that "play no role" in the deduction of the end-formula from the undischarged top-formulas – no matter in which particular way the deduction was performed –, and that can thus be properly removed viz. inserted without consequences for the fulfillment of this task – namely, the ones involved in the reductions described above. The soundness part of the normalisation thesis seems to amount to the claim that the removal viz. insertion of these patterns from viz. in a derivation is innocuous[19], in some sense, to its

19 Prawitz acknowledges that this may not be the case for the "redexes" of permutative reductions associated to the eliminations of disjunction and existential quantifier. But this does not stop him from putting the normalisation thesis forward anyway. For the sake of the argument, then, we shall henceforth not question

semantical value; and hence supposedly irrelevant to the determination of which proof is represented by the derivation itself in which it is eventually performed. The reasons given for such a claim are to be explored later (III.2.a). Uniqueness of normal form, in turn, adds to the normalisation procedure the flavour of an *evaluation*; one in which derivations are unambiguously judged with respect to which proof they represent, normal derivations playing the role of identity values – or, more in tune with the terms of the official formulation of the thesis, unique canonical representatives of *proofs*, which are the actual identity values. The situation can be regarded as analogous to what happens with e.g. numerical expressions in general and canonical numerals as representatives of natural numbers: the very fact that "3+2" ,"4+1", "1+2+2", etc. all ultimately reduce to "5" and to "5" alone can be regarded as suggestive of the fact that these numerical expressions have the same value, namely 5 – i.e. the "direct", disquotational value of "5". This particular kind of understanding of the normalisation procedure is presumably one of the main motivations behind one of the most popular reformulations of the normalisation thesis, used by e.g. Troelstra in his *non-extensional equality*, namely: Two proofs corresponding to deductions π and π' are the same iff π and π' reduce to the same normal form.

In what follows, an attempt will be made to investigate the extent to which the normalisation thesis makes justice to the informal motivational thesis formulated above by analysing its success in fulfilling certain important tasks necessary for it to adequately preserve the virtues of the informal ideas involved in it. But let us first try and describe certain basic semantical conceptions that seem to be involved in its formulation – namely, some regarding the idea of derivations as "representatives" of proofs –, so that we can try and properly evaluate the thesis in the intended respect. This will be done in a purposefully naive, as literature independent as possible fashion; in this way, we can sensibly nurture more hopes to avoid a viciously biased starting point for our discussion.

b. Derivations as "representatives" of proofs

Let us start by noticing some background semantical assumptions apparently involved in the formulation of the thesis, which are actually very important for, before anything else, the

---

the soundness of permutative reductions; nor will we question the soundness of so-called immediate simplifications. Instead, let us grant from the very outset that they are trivially identity preserving, so that they may be safely ignored without the need of adding provisos regarding them at the many places those would be due. The considerations regarding the normalisation conjecture in this section were thus made aiming basically at β- and η-conversions – and most especially at the first.

proper delimitation and comprehension of the problem at stake. The “official” formulation of the thesis carries with itself one crucial semantical departure point which constitutes, far beyond a mere characteristic determination of an answer to the question regarding identity of proofs, clear indications as how such question is supposed to be asked in the first place. It is the following: derivations are conceived as “representatives” of proofs, which are in turn conceived as their semantical values. Thus, the question to which the normalisation thesis offers an answer – namely, “When do two derivations represent the same proof?” –, taken as an interpretation of the question regarding identity of proofs, turns the latter into a quest for how to properly map certain linguistic entities – namely derivations – onto semantic values – namely proofs. Thus derivations are taken to *denote* – i.e. “represent” – proofs; the question regarding “identity of proofs” becomes indeed a question that regards, before anything else, *synonymy* or, more specifically, *co-referentiality* of derivations, and the identity of their semantic values, which we, following Prawitz, in this context call proofs – whatever these are considered to be – may in principle remain a simply unaddressed issue.

Now, once it is established that the relation between proofs and derivations is that the latter denote the former, some further conclusions can be drawn once we take into account some simple formal consequences of the normalisation thesis – at least if taken in its official formulation. The first to be noticed here is this: given that there are different equivalent derivations, it follows that the same proof may be denoted by different derivations. But what to say of the converse of this statement? – i.e. could a given derivation be taken to represent different proofs? It is important to notice that there is an asymmetry between these expressions of the two directions of the relation of denotation between derivations and proofs. It is due to an ambiguity of the expression “different proofs” in the formulation of the question concerning the second direction, which is absent in the case of the expression “different derivations”. While “different derivations” unequivocally refers in this context to any two distinct syntactical objects which we call derivations, different proofs may mean either, on the one hand, proofs to which the relation of identity of proofs, however we decide to understand it, cannot be correctly ascribed; or, on the other hand, proofs which are (say, numerically) distinct objects, regardless of whether or not these objects are identified by our relation of identity of proofs. If we take the expression in the first of these senses, it seems the question – namely, “could a given derivation be taken to represent different proofs?” – should then be answered negatively in the framework of the normalisation thesis: otherwise, derivations could denote *ambiguously*, i.e. a single derivation could have distinct and non-equivalent semantical values. And here we notice a further implicit semantical departure point of the thesis under discussion: the *non-ambiguity* of derivations with respect to their semantical value. On the second of these senses, however, one could answer to the question positively. To illustrate this, just take a derivation to be a syntactical *doppelgänger* of the proof it denotes; an exact "depiction" of it, so that there would be a distinct proof for every distinct derivation. This way, to see how a derivation could denote different proofs under the assumption of the normalisation thesis, it would suffice e.g. to say that the relation of identity viz. equivalence between the denoted proofs mirrors exactly that between the derivations that denote them; so that, say, two equivalent distinct derivations Π and Π' that denote respectively the consequently equivalent and distinct proofs **Π** and **Π'** would each also denote, respectively, **Π'** and **Π**. In this fashion, a derivation could denote many distinct – yet equivalent – proofs; indeed, as many as there are distinct derivations equivalent to it.

This second picture seems to distort a little the question "when do two derivations represent *the same* proof?" into "when do two derivations represent *equivalent* proofs?"; for although it serves as an answer to the first question, it seems much more in tune with the second. In any case, it has the virtue of doing more justice to the idea that the normalisation thesis answers to a question that actually concerns the identity viz. equivalence of *proofs*, rather than merely some arguably semantical equivalence between derivations; for it counts on a certain conception of proof that is a copy of that of derivation at the level of denoted objects. Thus, once equivalence between derivations is established to be in a certain way, so is identity between proofs. This semantical background to the study of properties and relations concerning proofs – and, in particular, to the problem of identity of proofs – clearly fits into the general strategy described by Kreisel in his 1971 Survey of Proof Theory II, p.111:

"The general nature of our problem is quite clear. Consider formal rules which are intended to formalize certain proofs; in other words, we have syntactic objects, derivations *d* which represent or describe mental acts ***d***, the proofs (which carry conviction);(...) the relation between *d* and ***d*** (...) is a particular case of the general relation between words and the thoughts they express. Since we are dealing with a "small" class of words, we can hope for more precise results than are known for the more general (and more familiar!) relation. (...)

Given a property P or a relation R between proofs our task is to find relations P_F, resp. R_F such that for all d of our formal system P_F (*d*) iff P(***d***) and R_F (*d, d'*) iff R(***d, d'***).

For exposition, we shall reverse this procedure, and first describe some formal relations P_F, R_F which will then be used to state the facts about the objects of principal interest, namely properties and relations of proofs."

The obvious problem with this "*doppelgänger*" approach is, of course, that it makes the denotational talk of "representation" involved in the formulation of the normalisation thesis become rather dispensable. The positing of proofs as objects distinct from and denoted by derivations becomes beard for Ockham's razor; one could, after all, simply take the normalisation thesis as the characterisation of a relation of equivalence of meaning of, say, an intensional character between derivations (which we could just as well call proofs), and regard the latter as themselves the proofs the identity of which we intend to investigate, rather than as syntactical or linguistic "representatives" of some objects of a more ineffable nature. Thus, one could completely eschew the denotational talk and formulate the thesis in the following terms: Two derivations/proofs are *synonymous* if and only if they are equivalent (in Prawitz's sense).

Nevertheless, the fact is that the denotational talk is there. So, to make better sense of it, it is easier to answer negatively to the question as to whether or not a derivation could denote different proofs also in the second of the senses mentioned[20], and understand the normalisation thesis as a solution to the question concerning identity of proofs in a quite simple way. Namely: first, as suggesting from the outset that each derivation, besides not being ambiguous, i.e. denoting *only equivalent* objects, denotes *only one* object viz. proof; second, as suggesting that the relation of equivalence between derivations is, as already observed, necessarily and sufficiently conditioned by their *co-referentiality* – i.e. by the fact that they denote *one and the same proof*; and third, as suggesting that the relation of identity

20 Although it hardly seems impossible to do otherwise. One could e.g. consider every distinct β-normal derivation Π_i of *A* from Γ as a canonical representative of a correspondently distinct proof $\boldsymbol{\Pi_i}$, and say that every derivation that reduces to Π_i would also denote proof $\boldsymbol{\Pi_i}$. Further, ler the $\boldsymbol{\Pi_i}$ be**,** despite distinct from one another, all *equivalent* proofs; say, by virtue of the fact of their correspondent canonical representatives being ultimately reducible to a unique βη-normal form. So, every derivation that reduces to Π_i would denote proof $\boldsymbol{\Pi_i}$, but not only; they would also denote all proofs equivalent to $\boldsymbol{\Pi_i}$. One could then make sense of such a situation in the following way: β-normal derivations are the *doppelgängers,* the canonical representatives of proofs conceived as the *intensional* contents of derivations, so that proofs that are β-convertible to one another share this intensional content. These intensional contents, the proofs, are *themselves* equivalent viz. "identical" when their canonical representatives reduce to a common βη-normal form. In such a picture, derivations may denote more than one proof; the proofs could not be simply considered Ockham-beard without further; and the treatment would yield an answer not only to a matter regarding semantical equivalence of derivations, but also identity of proofs in a literal way. Such a theory would nevertheless need much philosophical underpinning not to be merely an *ad hoc* way of making, simultaneously, the positing of proofs demanded by the denotational formulation of the normalisation thesis relevant; and the matter addressed by the normalisation thesis actually be, further than mere semantical equivalence of derivations, identity of proofs.

viz. equivalence concerning proofs themselves, the denoted objects, is to be understood as holding only between an object and itself, and is thus trivial and uninteresting. While this approach makes it harder to understand in which sense the normalisation thesis addresses a problem worthy of the label "identity of proofs" – instead of, say, co-referentiality of derivations –, it certainly accommodates much better the denotational mode of expression adopted in the formulation of the thesis.

Now, once it has been seen that it is not necessary at all to posit proofs as semantical denotations of syntactically conceived derivations in any sense or way in order to formulate the normalisation thesis in a sensible way; and that, as a matter of fact, positing such objects and denotational semantical relations make the thesis quite prone to be accused of displaying weaknesses such as Ockham-beardiness or lack of success in addressing a problem worthy of the title "identity of proofs"; one might then ask oneself just what would be the reasons to choose to formulate the normalisation thesis in its traditional, denotational fashion. Regarding this matter, I believe it is particularly interesting to stress that the normalisation thesis, as first put forward, is given *descriptive* rather than *definitional* contours. We shall see more precisely how this makes a difference to the matter under consideration shortly (III.2.a), when we shall dirty our hands with some observations more dependent on specialised literature. For now, let us just say this: the positing of proofs as independent objects denoted by derivations allows a quite Fregean-spirited solution to the question whether or not two derivations represent the same proof: just as the morning star and the evening star are identical by virtue of being the same celestial body or 117+136 and 253 are identical by virtue of being the same natural number – which implies that "the morning star" ["117+136"] and "the evening star" ["253"], in spite of their eventual difference in *Sinn*, are descriptions of the same object (i.e. have the same *Bedeutung*), and can thus be the arguments of a true and informative identity statement – so would, say

$$\begin{array}{c} \Pi_1 \quad \Pi_2 \\ \dfrac{\dfrac{A \qquad B}{A \wedge B}}{A} \\ \Pi_3 \end{array} \qquad \text{and} \qquad \begin{array}{c} \Pi_1 \\ A \\ \Pi_3 \end{array}$$

(and here I am using rather than mentioning the derivations) be identical by virtue of being the same proof – which implies that

$$“\ \begin{array}{c} \Pi_1 \quad \Pi_2 \\ \dfrac{\dfrac{A \quad B}{A \wedge B}}{A} \\ \Pi_3 \end{array}\ “ \quad \text{and} \quad “\ \begin{array}{c} \Pi_1 \\ A \\ \Pi_3 \end{array}\ “$$

(now I am mentioning the derivations), in spite of their eventually different “senses”, are representatives/descriptions of the same object, thus serving as arguments for a true identity statement. The denotational talk can thus be understood as an essential ingredient of the possibility of resorting to the idea that *proofs,* taken as *actual objects,* and their nature viz. the properties they display and relations in which they partake, are the ultimate foundation of an adequate criterion for the truth or falsity of statements concerning the semantical equivalence of derivations – which would otherwise be arguably solely founded on stipulations or conventions concerning the meaning (*Sinn,* or – to untie us from Fregean terminology and doctrine – meaning *intensionally* conceived) of the latter, ultimately rendering the normalisation thesis a purely definitional account of synonymy/identity of derivations/proofs. As already observed, though, the label “identity of proofs” for an investigation conceived in such terms is quite misleading: just as Frege’s investigation in *Über Sinn und Bedeutung* did not concern the identity of celestial bodies, natural numbers or other objects, but was rather “*über Sinn und Bedeutung”* – i.e. it concerned phenomena related to the semantics of the *linguistic expressions* that *denote* celestial bodies, natural numbers and other objects –, this one also seems not at all concerned with the identity of denoted objects, but rather with the semantical equivalence of certain linguistic expressions. Furthermore, there is still an elephant in the room if one wants to use this denotational approach to identity of proofs: while we know very well what celestial bodies are and how to identify them viz. tell them apart – and the same holds even for the more “abstract” natural numbers, especially when it comes to telling them apart (which is what matters the most for our present concerns) –, it does not seem that we have this much clarity about the elusive supposed objects we here refer to as proofs.

2. General reconstruction and criticism of the thesis

Regarding the informal idea presented in III.1.a as support to the normalisation thesis: Premisses (α) and (β) are, so to speak, the main matter of the driving informal thesis itself, for each of them corresponds, respectively, to the *soundness* and the *completeness* parts of the normalisation thesis, their formal counterparts. In order to make full justice to the importance of these two premisses, however, it is important to stress that the roles they play are not quite the same: (α) is the actual core of the doctrine, for it provides the positive criterion according to which proofs should be identified, while (β), in stating simply that nothing that is not identified by (α) should be identified, is of a dependent and merely negative nature.

At the very informal and undetailed level they have been presented here, (α) and (β) seem remarkably resistant to disagreement; indeed almost trivial. This stems from the fact that they are built upon an appropriate notion of *irrelevance*; one such that its proper comprehension makes it a truism that irrelevant differences are always identity preserving, and also that any difference that is not irrelevant is also not identity preserving.

In the way down to the concretion of their formal counterparts, however, their resistance to disagreement diminishes considerably. Premiss (α), in the first place, can be regarded as being preliminarily specified to a still informal restriction of itself, of which the soundness part of the thesis is intended as a mere *formalisation*. Let us call this informal particularization (α1), and leave undetermined the question of what more precisely it should look like. Despite the fact that it is merely a particular case of (α) – actually, precisely *because* it is merely a particular case of (α) –, (α1) brings about a severe problem when combined with (β); one of the most severe problems, indeed, that the normalisation thesis faces as it stands, namely: the difficulty of supporting its claim to completeness.

It is thus presumably having something that fits into the role of (α1) – i.e. some form of restricted version or specification of (α)[21] – in mind that advocates of the normalisation thesis end up simultaneously, on the one hand, depositing their full confidence upon its alleged formal counterpart – namely, the *soundness* part of the normalisation thesis –, to the point of deeming it obvious (see Prawitz 1971, Kreisel 1971 (SPT II); cf. Feferman's 1975 review of Prawitz 1971); and, on the other hand, acknowledging (β), the *completeness* part of the normalisation thesis, as a sort of Achilles' heel of the proposal – an attitude rather more frequent previously to the obtention of the Post-completeness/maximality results concerning

21 Just to provide the reader with an idea of what (α1) could roughly look like, we could suggest something in the spirit of the following formulation: any two proofs that can be mechanically transformed in one another by sequences of additions and/or removals of irrelevant features are not significantly different. This is, again, rough and certainly not the best possible characterisation of (α1), but it will do to convey those eventually bothered by the lack of an image an idea of what such a premiss may involve. How (α1) is to be properly formulated is of no consequence to the sequel.

the notion of identity of proofs yielded by the normalisation thesis, widely accepted as a stamp of approval of the *completeness* part by enthusiasts (we shall return to the issue of the maximality results soon, in III.2.b). Indeed, in order to accept the soundness of (α) – which, as observed, is difficult not to do –, it seems inevitable that one also accepts the soundness of a specific case of it. As much as we keep with a completely trustworthy claim to soundness in passing from (α) to (α1), however, this does not prevent at all the loss of justification to a claim to completeness in the second case. And the reason is the very same that supports the trust in the soundness of (α1): it is merely a specific case of (α), the soundness of which we already accepted, and may thus in principle leave out some proofs that (α) could identify.

The soundness part of the thesis, however, is quite distinct from both (α) and (α1), regardless of how the latter is determined:

(αF) If two derivations are (βη-)equivalent, then they represent the same proof.

It seems that(αF) inspires far less confidence than their informal counterparts, simple and solid objections to its soundness having been presented in the literature (see. e.g. Troelstra 1975, Feferman 1975 review of Prawitz 1971, Došen 2003), two of which shall be scrutinised in the sequel (in III.2.a.1 and III.2.a.2). We shall nevertheless first try to reconstruct (αF) as a formal specification of (α1) and understand the reasoning behind its formulation.

a. The case for the soundness of (αF) or Prawitz's teleology of proofs.

In his 1965 doctoral dissertation, p.33, Prawitz describes a relation observed to hold between the usual introduction and elimination rules of usual logical constants, which he, following Lorenzen's terminology, calls the *inversion principle*:

"*Let α be an application of an elimination rule that has B as consequence. Then, deductions that satisfy the sufficient condition (…) for deriving the major premiss of α, when combined with deductions of the minor premisses of α (if any), already "contain" a deduction of B; the deduction of B is thus obtainable directly from the given deductions without the addition of α."*

The principle refers first to the following deductive pattern:

$$\dfrac{\dfrac{\Pi_1, \ldots, \Pi_n}{*^n(A_1, \ldots, A_{n_1})}\, {*^n}\text{-}introduction \qquad \begin{matrix}\Gamma_1 & & \Gamma_m \\ \vdots & \ldots & \vdots \\ C_1 & & C_m\end{matrix}}{B}\,\alpha$$

Its claim is roughly that B, the conclusion, could always be obtained by eschewing the introduction of $*^n(A_1, \ldots, A_n)$ in the following way:

$$\begin{matrix}\Gamma_i \\ \vdots \\ C_i \\ \vdots \\ \Pi_j \\ \vdots \\ B\end{matrix}$$

Even though this is evidently not the case in general, the restriction to the intended constants and rules indeed confirms the principle. For each such constant, the former deductive pattern can be re-written as the latter according to the reduction steps described by Prawitz in the sequel (pp.36 – 38) – the same mentioned here some sections before.

Notwithstanding, it is worth noticing that, before the presentation of the reduction steps, Prawitz rephrases the sober observations that constitute the formulation of the inversion principle as a somewhat cautiously – yet still quite loudly – expressed semantical battle cry, which occurs in passages such as the following (emphasis by myself):

“The inversion principle says in effect that <u>nothing is ‘gained’</u> by inferring a formula through introduction for use as a major premiss in an elimination.” (pp. 33 - 34)

“We note that a consequence of an I-rule which is also major premiss of an E-rule constitutes <u>a complication in a deduction</u>. As such a complication can be removed (…), we may ask whether it is possible to transform every deduction to a corresponding ‘normal’ one which proceeds, so to say, directly, without any <u>detours</u>, from the assumptions to the end-formula.” (p.34)

He still adds, in the very same 1971 paper where he would, later, put the normalisation thesis forward:

“Here, I shall consider a more direct way of making the inversion principle precise. Since it says that <u>nothing new is obtained</u> by an elimination immediately following an introduction (of the major premiss of the elimination), it suggests that <u>such sequences of inferences can be dispensed with.</u>" (p.247)

And right after presenting the thesis, he claims that the previous discussion of the inversion principle should make (αF), i.e. its soundness part, somehow obvious.

Notice that β-redexes viz. maximum occurrences of formulas are here already most definitely characterised as somehow superfluous. But superfluous to what purpose?; according to which criterion? (see the discussion in 4.2 on the suppression of the position of criterion that marks the contemporary appropriation of the relation of synonymy, as opposed to that of Aristotle). Unless we are able to answer this, accepting or rejecting such a characterisation shall remain equally arbitrary options. And the answer apparently suggested in the passages themselves is this: superfluous to the purpose of obtaining the end-formula of a given derivation from its assumptions.

In the first place, the fact should be acknowledged that β-redexes are indeed unnecessary to the derivation of any given end-formula A from any given set of assumptions Γ; and indeed completely removable from any given derivation of a given A from a given Γ in a mechanical way, without prejudice to the status of derivation of the formula-tree in question at any stage of the procedure – this much is guaranteed by the normalisation theorem, proved by Prawitz himself. Now, this helps us understand the purpose to which β-redexes are superfluous viz. add nothing new, etc., namely: to the construction of any specific derivation of a formula A from a set of assumptions Γ. They neither have the power to prevent, if suppressed, nor to favour, if added, that A is in fact derived from Γ in a given derivation. As the structure of the reductions proposed by Prawitz makes clear, the presence of a β-redex in a derivation of a formula A from a set of assumptions Γ amounts to the addition of a formula which is either redundant, i.e. literally an unnecessary repetition within the derivation, or simply *irrelevant*[22] to the obtainment of the end-formula from the top ones.[23]

But between the acknowledgement of this much and the conclusion that addition or suppression of β-redexes preserve the identity value of a derivation lies an abyss, upon which no bridge seems to stand (see the sections below on objections to the soundness of the normalisation thesis). Unless, of course, one is ready to admit a most peculiar kind of teleology of derivations and proofs; a metaphysical doctrine according to which the "final cause", so to speak, the purpose of a proof is the obtainment of its (end)result from the assumptions; and that the way to do this by means of a derivation is obtaining the end-formula, the formal counterpart of the (end)result, from the top-formulas, the formal

22 It is important to stress the difference between the use of the terms "unnecessary" and "irrelevant" in this context. While many normal derivations could be said to have occurrences of formulas that are, strictly speaking, unnecessary to the obtainment of its end-formula from its undischarged top ones, none of them could be said to contain occurrences of formulas that are *irrelevant* to the same end. Consider, for example, the following two normal derivations of *A* from $(A \wedge B) \wedge (B \wedge A)$:

$$\dfrac{\dfrac{(A \wedge B) \wedge (B \wedge A)}{A \wedge B}}{A} \qquad \dfrac{\dfrac{(A \wedge B) \wedge (B \wedge A)}{B \wedge A}}{A}$$

Notice that there is a clear sense in which the occurrences of $(A \wedge B)$ in the derivation to the left and of $(B \wedge A)$ in the one to the right are unnecessary: the very existence of a normal derivation of *A* from $(A \wedge B) \wedge (B \wedge A)$ which does not contain one or the other makes evident what is meant by that. It is nevertheless essential that $(A \wedge B)$ (resp. $(B \wedge A)$) occurs in the derivation to the left (resp. right) for the obtainment of the end-formula from the undischarged top ones, in the sense that its *suppression* would in any case tamper with the status of derivation of the formula-tree viz. turn it into an "ill-formed" derivation. The qualification of the mentioned repetitions as unnecessary thus entails neither that all nor that some specific occurrence of the repeated formula is unnecessary – only the repetition itself is unnecessary according to the present manipulation of this expression.

23 It is worth noticing that a reduction proposed by Ekman in his doctoral thesis seems to address the very same issue. He proposes the following reduction scheme, where the derivation to the left, Δ, reduces to the one to the right, Δ', and where all the undischarged top-formulas of Π are also undischarged in Δ:

$$\begin{array}{c} \Pi \\ A \\ \Sigma \\ A \end{array} \quad \Rightarrow \quad \begin{array}{c} \Pi \\ A \end{array}$$

As one clearly sees, this reduction also avoids repetitions in a derivation which are utterly unnecessary to the obtainment of the end-formula from the top ones; and in so doing it might also erase formulas in the derivation the occurrence of which was not relevant to the same purpose.

There are nevertheless some crucial differences: in the first place, while Prawitz's reductions aim at eliminating irrelevant formula occurences that might by chance also be repetitions, the point of Ekman's reduction is to eliminate repetitions, which may eventually also cause the elimination of irrelevant formulas which are not repetitions. Secondly, Prawitz's reductions are motivated by the inversion principle, which holds specifically for the constants he deals with, while the motivation for Ekman's reductions is in principle independent of particularities of the constants in the deductive system. Furthermore, Ekman's reduction yields a trivial notion of identity of proofs of a given *A* from a given *Γ,* which is not the case of Prawitz's reductions. Lastly, Prawitz's redexes seem to be properly understandable as superfluous according to his characterisation both in a global and in a local sense with respect to the derivations where they occur, while Ekman's redexes only seem to be properly understandable as superfluous in general in a global sense.

counterparts of the assumptions. The suggestion that precisely such a teleology is assumed by Prawitz is actually corroborated by himself more than once. Answering to criticism raised by Feferman to his notion of identity of proofs, for instance, he says that he looks upon proofs not “as a collection of sentences but as the result of applying certain operations to obtain a certain end result” (Prawitz 1981, p.249). In an answer to an article written by Sundholm addressing his views on the proof-act vs. proof-object distinction, in turn, he gives a most eloquent indication of this in quite general, analogical terms, which I will quote *in extenso*:

“(...)it is certainly undeniable that a word like building can be used for the act that an agent performs when he builds something and also for the building, a house for instance, that is the result of his act. It is equally obvious that to acts like constructing, travelling, cooking, and planning there are associated certain objects, in this case referred to by other words than those used for the act, viz. constructions, travels, dishes, and plans, which are results of the performances of the acts. Let us call the object that in this way results from an act an act product. (…)

Finally we may speak about the goal or the purpose of an act, naturally also called the object of the act, and which, again following Sundholm, I shall refer to by this term hoping that it shall not cause confusion. The long-range goal of an act varies of course but there may be goals that are conceptually connected with the act. Often the act product is just the goal or object of an act, or to be careful we should perhaps say that it is not the house but there being a house which is the goal of building. The object of an act may perhaps also be counted as an outcome of the act, but it may sometimes be of interest to differentiate between the object of the act and the act product. Travelling may sometimes have just the resulting journey as its object, but the object of travelling to Italy is more likely to be the arrival to Italy; of course, one could say that in the latter case the journey is not the act product but the act process. Planning is a clearer example, the planning of a conference, say. The act product is the plan of the conference, which may contain such things as a program, a time schedule, a budget etc. But the planning is not made for its own sake: the conference, or the conference taking place, is the goal, the object of the planning. The act process on the other hand may contain such elements as the kind of meetings held, the program committee, and other aspects of how the planning was made.

It is now of interest to see how these concepts can be applied to acts such as assertions, inferences, observations, verifications, and proving.” (pp. 319 – 321 of Prawitz, D.

Comments on the papers, Theoria 64 (Special issue on the philosophy of Dag Prawitz), pp 283-337, 1998.)

The adoption of such a peculiar teleological metaphysics of proofs is nevertheless not to be regarded as the only root of the confidence in the soundness part of the normalisation thesis. It makes, one could say, the context become clear in which the idea of β-redexes viz. introductions of maximum formulas as irrelevant deductive patterns in a proof – "detours", as they are so often called – sounds like the triviality its proponents seem to consider it to be; i.e. it exposes the criterion according to which these deductive patterns become *intensionally* innocuous. However, there is also a more extensional, *denotationally* conceived underpinning to the faith in the soundness part of the thesis – one that also gives more substance to the denotational terms in which it is formulated. It has to do with a certain approach to proof-theoretic semantics championed by Prawitz and Dummett, centered in explaining the meaning of propositions and logical constants by means of "what counts as (canonical) proofs of them"; viz. by means of their *introduction* in proofs. This doctrine is closely related to the so called Brouwer-Heyting-Kolmogorov (BHK) interpretation of the logical constants, and might be reasonably regarded as favouring a view of proofs as, essentially and before anything else, "*meaning-makers*", so to speak, for respective logical constants, and, analogously, "truth-makers" for respective propositions/assertions. Thus, to illustrate how this doctrine provides a basis for the soundness of the normalisation thesis, let us take proofs to be objects – linguistic, mental or whatever – characterised by the BHK clauses, and natural deduction derivations to be linguistic entities that have meaning by virtue of the fact that they denote them. When one accepts beforehand such a view of what proofs are as that expressed by the BHK clauses, it then becomes quite tempting to see normal viz. *canonical* derivations as direct, "disquotational" representatives of proofs; for the structure of these derivations *mirrors*, so to speak, the structure of the denoted proof. The reason is that BHK-clauses characterise proofs in a recursive way as proofs of a proposition of a given form built out of proofs of other propositions by means of operations analogous to the respective introduction rules of natural deduction; and this makes their structure match the structure of (inductively conceived) normal derivations of formulas which correspond to the propositions of relevant form. For example, the BHK clause for conjunction states that a proof of $A \wedge B$ is a pair constituted by two proofs, one of *A* and one of *B*; analogously, as shown by Prawitz already in his 1965 dissertation, a *normal* derivation of a conjunction is obtained by ultimately joining together, by means of the introduction rule for conjunction, two derivations, one for each of the two conjuncts (see Prawitz 1971 *Ideas and Results*, section II.3.2 on the form of normal derivations; especially corollary 3.2.4.1).

However, since derivations are inductively conceived rather than recursively as in the case of proofs, and since one may apply elimination rules to build them, there can be non-normal derivations. Thus, unsurprisingly, when it comes to non-normal derivations, the mirroring of the structure of proofs is lost; and with it, the solid reason to regard these derivations as representatives of proofs. The normalisation theorem, however, offers a way to amend this: it shows not only *that* but also *how* one can mechanically "reorganise", so to speak, a non-normal derivation so as to turn it into a normal one. If one pays close attention to the structure of the reductions proposed by Prawitz to remove maximum formulas (segments), they are ways of avoiding an application of an elimination rule to obtain a given conclusion from a given (major) premiss within a derivation, *provided that this premiss was itself first obtained by an introduction rule* – i.e. it is, put roughly, a way to enable the obtention of the mentioned conclusion without resorting to elimination, provided that its premiss was also so obtained. Thus, once the reductions are accepted, a non-normal derivation can be seen as a *method* to obtain a normal derivation – just as the numerical expression "117+136" can be seen as a method to obtain the numeral "253". So, just as "117+136" and "253" denote the same number, namely 253, a (non-normal) derivation and the normal derivation to which it reduces – which, as already observed, is unique – are taken to denote the very same BHK-proof; the latter directly, the former "indirectly". And, in general, derivations one of which reduces to another denote the same proof.

In such a theoretical framework, in which proofs are recursively conceived acording to the proposition they prove as dictated by the BHK-clauses, it does not at all come as a surprise that the point of a proof is to "build" a specific given proposition according to the meanings of the constants involved in it – the meanings of which are in turn also determined by what their introduction conditions determine to count as proofs of propositions in which they are the main connectives.

Furthermore, it is important to notice that the normalisation thesis was proposed by Prawitz as a *conjecture:* i.e. a possibly true, possibly false description of the conditions under which derivations are semantically equivalent. Thus, it is somehow assumed that the aimed relation of semantical equivalence between derivations is, in some sense, founded upon some independently given object(s) of investigation, to which one could resort in order to decide whether descriptions of it are true or false. Now, it is a clear fact that the normalisation thesis lacks any intention of describing how derivations are *in practice* judged with respect to their eventual semantical equivalence – so real life is not fit to play the role of criterion for the truth/adequacy of the thesis. But what would then be? It is easy to see that if one takes the thesis as a purely intensional, non-denotational account of semantical equivalence between derivations in terms of the formal equivalence relation specified by Prawitz, this quite inevitably deprives the thesis of the possibility of addressing any *relevant* "truth-making" reality; rather than a possibly true, possibly false *description* of the conditions under which derivations are semantically equivalent, it would thus be a *definition* of this much. This is of course not to claim that such a definition would necessarily be an arbitrary stipulation; there could be many ways to argue quite respectably for e.g. the idea that Prawitz's reductions are meaning-preserving. But ultimately, a notion of semantical equivalence so intensionally conceived could not be supported by any facts if not by what being semantically equivalent eventually *means* for derivations – i.e. a matter of definition, not of fact. On the other hand, when derivations are taken as representatives/descriptions of BHK-*proofs,* the latter in turn taken as independent objects, one can appeal to the fact that *these very proofs* are themselves such that certain (different) derivations – which might even perhaps differ in meaning in some other respect – will have viz. share them as their denotation.

Thus, in a single strike, the mystery around the teleology disolves – well, at least it is swept under the carpet: the adoption of a BHK-like view on proofs from the outset of the investigation remains simply unjustified[24] –; and the conjectural character and denotational terms in which the thesis is originally formulated find, together, a reason to be.

Apart from the many and solid direct objections one might have against this view of proofs as meaning- viz. truth-makers and the faithfulness of the particular formalisation of it involved in the formulation of the normalisation thesis, there is a disturbing issue regarding the (lack of) coherence of this particular approach to proof-theoretic semantics and Prawitz's own philosophical agenda of a *general proof theory*. According to the author himself, general

24 One would do well also to remember that the acceptance of such a view on proofs from the outset of the investigation breaks one of the guidelines suggested in the previous chapter as essential to the obtainment of clarity about how identity of proofs may contribute to our understanding of what a proof is, namely: one should not have to rely on a previously fixed conception of what a proof is (BHK in particular) in order to adequately account for the question of identity of proofs.

proof theory is distinguished by the fact that it is a study of proofs “in their own right”. He states that “In *general proof theory*, we are interested in the very notion of proof and its properties and impose no restriction on the methods that may be used in the study of this notion.” (ideas and results 1971, p.236) – and arguably, as e.g. Kosta Došen puts it, identity of proofs is at the heart of such a study, for only by means of an adequate account of it can we hope to answer satisfactorily the driving question “what is a proof?”. But studying proofs inasmuch as they are conceived beforehand as being, essentially and primarily, the semantic foundations for the explanation of the meanings of *other entities*, such as logical constants and propositions, is hardly to consider them “in their own right”; rather, their meaning is thus reduced and distorted in order to allow for attempts at satisfactorily explaining the meanings of other things – things which are undoubtedly relevant and interesting, but the construction of which does most certainly *not* exhaust even the central interesting aspects of proofs and their potential; not even those within the competence of logic. Furthermore, it is difficult to see how an investigation of identity of proofs so conceived could help in any significant way to understand what a proof is in a literal and relevant enough sense – after all, the whole enterprise, as well as the solution it yields, seem to depend quite heavily on a preconceived and quite specific idea of what a proof is. I daresay that, while we keep determined not to give up a certain way of explaining the meaning of logical constants and propositions by means of proofs, our approach to the notion of proof will at best always be viciously biased by a semantical goal we force proofs to fulfill from the very outset, and our explanation of their identity and nature will thus result unavoidably *ad hoc* – or, maybe even worse, no more than a *petitio principii.*

a.1. An objection to the soundness of (αF) suggested by Kosta Došen

In the same text where he suggests the notion of propositional identity (viz. synonymy) discussed above, Kosta Došen voices a possible objection to the normalisation thesis. His point is that one might doubt the putatively “reliable” direction of the thesis, i.e. its claim to soundness, based on the observance of the fact that there are certain βη-equivalent derivations which generalise differently viz. to derivations of different formula-schemes. He presents two typed lambda-terms as example, which correspond to the following two derivations:

$$\dfrac{\dfrac{\dfrac{[A\supset A]^1 \quad [A\supset A]^2}{(A\supset A)\wedge(A\supset A)}}{(A\supset A)\supset((A\supset A)\wedge(A\supset A))}{\scriptstyle 1,2} \qquad \dfrac{[A]^3}{A\supset A}{\scriptstyle 3}}{(A\supset A)\wedge(A\supset A)} \quad \gg_{\beta\eta} \quad \dfrac{\dfrac{[A]^1}{A\supset A}{\scriptstyle 1} \qquad \dfrac{[A]^2}{A\supset A}{\scriptstyle 2}}{(A\supset A)\wedge(A\supset A)}$$

Indeed, while the first derivation cannot be further generalised, the second, which one obtains by removing the β-redex/maximum formula that occurs in the first – and is therefore βη-equivalent to it – clearly generalises to a derivation of $(A\supset A)\wedge(B\supset B)$. As Došen himself points out, this phenomenon does not depend at all on the introduction of conjunction (since he discusses lambda-terms, he speaks of surjective pairing), and may arise within purely implicational logic (viz. merely with functional types). The following pairs of (βη-)equivalent derivations illustrate this point adequately:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[A]^2}{C\supset A} \qquad \dfrac{[B\supset C]^3 \quad [B]^1}{C}}{A}}{B\supset A}{\scriptstyle 1}}{A\supset(B\supset A)}{\scriptstyle 2}}{(B\supset C)\supset(A\supset(B\supset A))}{\scriptstyle 3} \quad \gg_{\beta} \quad \dfrac{\dfrac{\dfrac{[A]^1}{B\supset A}}{A\supset(B\supset A)}{\scriptstyle 1}}{(B\supset C)\supset(A\supset(B\supset A))}$$

In this example, while the first derivation cannot be further generalised, the second clearly generalises to a derivation of $C\supset(A\supset(B\supset A))$.

$$\dfrac{\dfrac{\dfrac{[A\supset B]^2 \quad [A]^1}{B}}{A\supset B}{\scriptstyle 1}}{(A\supset B)\supset(A\supset B)}{\scriptstyle 2} \quad \ll_{\eta} \quad \dfrac{[A\supset B]^1}{(A\supset B)\supset(A\supset B)}{\scriptstyle 1}$$

Similarly, in the example above, while the first derivation cannot be further generalised, the second generalises to a derivation of $A \supset A$.

Notice further that the last two examples above together show that neither β- nor η-conversion preserve the generality of a derivation in general; not even when we are restricted to the minimal implicational fragment of propositional logic. Both β and η reductions may obliterate (or create) bonds of variables so that the form of some formula occurrences becomes superfluous (or essential) in the reduced derivation, even though they were essential resp. superfluous before the reduction. Indeed, such obliterations are what leads to the variation in the generality, as observed in the cases just displayed.

a.1.1 Interpretation and radicalisation of Došen's objection

The idea behind Došen's objection – viz., that two derivations that generalise differently should be considered different – is one of the pillars of the generality thesis on identity of proofs, addressed by the author in the same text. Interestingly enough, when Prawitz first presented the normalisation thesis, he seems to have resorted to the very same idea to give some support to the proposal. He presented two different normal derivations – that are therefore not βη-equivalent and hence *ex hypothesi* not identical – of the theorem $(A \supset A) \supset (B \supset (A \supset A))$, one of which generalises to a derivation of $C \supset (B \supset (A \supset A))$, the other to a derivation of $A \supset (B \supset A)$, and claimed that they are clearly based on different "proof ideas" – the elusive concept which plays the role of Holy Grail in the philosophical discussion concerning identity of proofs which takes place in the literature.

Now, the reason to find it odd that two derivations that generalise differently in the way the ones above do should be identified/considered equivalent lies on two main assumptions: (i) the widespread practice in the literature on identity of proofs of only considering criteria of identity belonging to the section "restricted to the result" of the taxonomy proposed here in the previous chapter; and (ii) the interpretation of the generalisation procedure as somehow making explicit the "real result" (see. e.g. Widebäck 2001, p.53) proved by means of a derivation. In short: the possibility of two proofs of different results being equivalent is most usually ruled out from the very outset, and having generalisations with different end-formulas means roughly proving different results. This is a way to explain why the examples presented above should be understood as posing a challenge to the normalisation thesis.

In my opinion, this argument against the normalisation thesis is underdeveloped as it stands. A light flare of open-mindedness is enough to safeguard the advocate of the thesis from the objection: it would suffice them to e.g. drop the idea that the normalisation criterion belongs to the section “restricted to the result” of the taxonomy, i.e., accepting that there is no problem in principle in identifying proofs of different results – which amounts to simply rejecting (i).[25] Notwithstanding, some formal facts closely related to the one observed by Kosta Došen allow one to ground scepticism towards the normalisation thesis in a more robust way.

THEOREM: Some theorems cannot be obtained as the conclusion of a maximally generalised closed normal derivation.

In other words, there are theorems which express no “real result” proved by a normal derivation.

PROOF: let us consider the theorem $A \supset (A \supset A)$. It is easy to verify that this theorem has only two closed normal derivations, namely:

25 Another option also available is, obviously, rejecting (β). This could be done by, say, sticking to the idea that the actual end- and undischarged top-formulas of a derivation are perfectly faithful expressions of the result proved by means of such derivation. But that seems to yield a notion of proof that is too narrow from the very outset – in case one maintains (α), for instance, one could not even identify two proofs one of which is a proper schematic instance of the other. (β) could of course also be rejected by other means – one, suggested to me by Prof. Schroeder-Heister, amounts to considering that what the generalisation procedure reveals is not to be understood as the “real result” of the proof carried out by means of a derivation, but rather as the “potential” of such a proof – where the potential of a proof is the most general result the computation carried out to perform the proof may prove. In such a case, though, it is rather unclear what one should consider to be the actual result of a proof: in case it is the end formula of the derivation by means of which the proof was performed, we are back to the previous case, i.e. the concept of proof seems too narrow; in case it is however the same as the proof’s potential, then there is no real denial of (β). Distinguishing any shade of grey between these two options seems difficult, and picking one, if any, would be just arbitrary – and, as we shall see in the sequel, more extreme options in the direction of the second one involve significant problems.

$$\dfrac{\dfrac{[A]^1}{A \supset A}\,1}{A \supset (A \supset A)} \qquad \dfrac{\dfrac{[A]^1}{A \supset A}}{A \supset (A \supset A)}\,1$$

Clearly, though, neither of the derivations above is maximally generalised: the first can be further generalised to a derivation of $B \supset (A \supset A)$, and the second, to one of $A \supset (B \supset A)$.

Q.E.D.

Even if we are under the assumption of the normalisation thesis, this result alone merely states that there are some theorems which cannot be derived *normally* as "real results" of some proof. At this point, it could still well be the case that such theorems cannot be derived *at all* as real results of some proof, i.e. as conclusions of a maximally generalised derivation – in other words, that such theorems are not real results obtained in proofs at all, but rather mere "instances", so to speak, of what proofs actually prove. However, considering the same theorem used as example above, it can be shown that this hypothesis does not hold in general:

THEOREM: There are theorems that cannot be derived as conclusions of maximally generalised closed normal derivations but that can be derived as conclusions of non-normal maximally generalised closed derivations.

PROOF: As shown above, the theorem $A \supset (A \supset A)$ cannot be derived as conclusion of a maximally generalised closed normal derivation. It is nevertheless the conclusion of the following maximally generalised non-normal closed derivation:

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[A]^1}{A \supset A}}{A \supset (A \supset A)}{\scriptstyle 1} \quad \dfrac{[A \supset A]^4 \quad [A]^2}{A}}{A \supset A} \qquad \dfrac{[A \supset A]^4 \quad [A]^3}{A}}{A}}{A \supset A}{\scriptstyle 2}}{A \supset (A \supset A)}{\scriptstyle 3}}{(A \supset A) \supset (A \supset (A \supset A))}{\scriptstyle 4} \qquad \dfrac{[A]^5}{A \supset A}{\scriptstyle 5}}{A \supset (A \supset A)}
$$

Q.E.D.

A possibility closely related to the one the general validity of which was just refuted by the example above is indeed rhetorically entertained by Hindley in his 1997 *Basic Simple Type Theory,* p.93; there, the author wonders about an eventual “aristocracy” of implicational theorems that are principal types of some closed lambda-term (i.e., by Curry-Howard, theorems that are conclusions of some maximally generalised closed derivation), in opposition to an eventual class of implicational theorems that are principal types of no lambda-term. He rules this possibility out in the immediate sequel, however, by presenting the so called converse principal type theorem, which in the present context could be given the following formulation:

THEOREM (CPT theorem): every implicational theorem can be obtained as the conclusion of some maximally generalised closed derivation.

PROOF: See Hindley’s presentation in *Basic Simple Type Theory* (from p.93). Q.E.D.

(CPT theorem, together with the normalisation theorem, directly yields a corollary:

COROLLARY (weak “principalisation”): Every implicational theorem *A* has a derivation Π that can be turned into a maximally generalised derivation Π’ of *A* exclusively by means of β- and η-conversions.

PROOF: Direct consequence of CPT and normalisation. Q.E.D.

Given this, one might find it worth pursuing a proof of what would be a stronger result:

Conjecture ("principalisation"): Every derivation Π of an implicational theorem *A* can be turned into a maximally generalised derivation Π' of *A* exclusively by means of β- and η-conversions.

Or maybe an even stronger:

Conjecture (strong "principalisation"): Every sequence of βη-reduction of every derivation Π of an implicational theorem *A* contains a maximally generalised derivation Π' of *A*.

The corollary itself is at best somewhat suggestive, but any of these conjectures, if proved, should show an interesting connection between β- and η- conversions and the property of maximal generality of a derivation. This could indeed ground a reinterpretation of the notion of normality of derivations to be attached to such conversions, alternative to that of Prawitz. One could thus understand normality and normal forms of derivations as having to do with bonds of variables and maximal generality, the alternative guiding notions, instead of subformula principle and absence of "detours" – to be normal would then mean to be a derivation of its own principal type, rather than a detour-free derivation of its conclusion.)

Now, the combination of the first with any of the two latter results just presented – no matter if the stronger or the weaker, which follows from it – provides a good argument against the normalisation thesis. If one takes seriously the idea that principal types – i.e. results of maximally generalised derivations – are good formal counterparts of the results of proofs carried out by means of their respective derivations (or, at any rate, that they are better formal counterparts of these than end-formulas taken crudely, which seems plausible), then these results entail that there are some results of proofs that simply cannot be proved normally viz. canonically.

Were the problem merely the existence of proofs of different results in a same class of equivalence, as Došen's point seems to suggest, then it could in principle still be the case that the results of non-normal derivations "altered" by means of the normalisation procedure could be obtained by means of some other canonical derivation than that to which the non-normal derivation in question reduces. Furthermore, as pointed above, the advocate of the normalisation thesis could still simply relax its original requirements and accept that proofs of different results can be identified. But the problem is actually more serious: since there are some real results which can only be obtained non-canonically, the notion of canonical proof that underlies and motivates philosophically the normalisation thesis is incomplete with respect to the results that can be proved. This means that, for the sake of the completeness of the deductive system with respect to what we have understood as the actual results of proofs, we simply cannot do without non-normal viz. non-canonical derivations.

From a merely formal standpoint, this situation indeed implies the existence of an “aristocracy” of theorems; though one of a different kind from that hypothesised by Hindley. This would consist in the set of theorems which can be proved as conclusions of maximally generalised *normal* derivations, in opposition to those which cannot, i.e. those which can be proved normally only if not as a conclusion of a maximally generalised closed derivation (which are all others, by the normalisation theorem), those which can be proved as a conclusion of a maximally generalised closed derivation only if not normally (which encompass at least all other implicational ones, by the CPT theorem presented above), and eventually also those which cannot be proved as a conclusion of a maximally generalised closed derivation at all (if any; I do not know whether or not there are such theorems).

How to interpret the meaning of the existence of such an aristocracy is of course a debatable matter. One could, for instance, take the membership in the aristocracy as the proper criterion to understand what a real result of a proof is, and indeed return to an understanding of the normalisation thesis as belonging to its original category in the taxonomy, viz. restricted to the result. This would nevertheless bring about the rather unusual consequence that the actual result proved by, say, a maximally generalised derivation of $A \supset (A \supset A)$ is in fact expressed by either $A \supset (B \supset A)$ or $B \supset (A \supset A)$, which are the conclusions of the correspondent possible maximally generalised normal forms of such a derivation – and that, in general, some theorems, including some that are conclusions of maximally generalised closed derivations, do not really express proof results.

But still, even if one does not go this way and prefers accepting that there can be equal proofs of different results, the semantical significance of the existence of such an aristocracy within the context of the Prawitz-Dummett proof-theoretic semantics that underpins the normalisation thesis is quite heavy: given that the meaning of a proposition is established by what counts as a canonical proof of it, and that a proposition expressed by a theorem not belonging to the aristocracy – e.g. $A \supset (A \supset A)$ – has no such proof (the fact that its correspondent theorem has such a *derivation* being here irrelevant, for, given the assumptions under which we now are, this theorem is not the proper expression of the result of the *proof* carried out by means of such a derivation), its meaning (say, as a valid proposition) is at the very best inevitably subordinated to, or dependent on the meaning of the aristocrats with which it is related, namely the conclusions of the maximal generalisations of the normal forms of the maximally generalised derivations of which it is a conclusion – in the case of the example given, $A \supset (B \supset A)$ and $B \supset (A \supset A)$. Given the fact that the deductive construction of $A \supset (A \supset A)$ – and of at least any implicational theorem whatsoever, for that matter – as the expression of the result of a proof which by no means demands – or even allows – more general conclusions is perfectly possible, it is very odd to tie oneself to a semantics that implies a dependence of the proof-theoretic meaning of the non-aristocrats on that of the aristocrats merely for the sake of sustaining that β- and η- conversions are identity preserving. In a sentence: there is hardly any reason that is not essentially *ad hoc* why normality (understood as freedom of "detours") should be seen as playing a more prominent role than principality (understood as maximal generalisation of a derivation with preservation of the structure of inference rules and assumption discharges) in the proper comprehension of identity of proofs .

It is therefore not merely because reductions may alter the result proved by means of a derivation, but rather because only by means of non-normal derivations can we express proofs of certain results – a rather different formal fact – that it seems really strange to believe that the differences between normal and non-normal forms of derivations in general are innocuous in any sense relevant to the discussion of identity of proofs.

a.2. Feferman's objection to the normalisation thesis

In his 1975 review of Prawitz 1971 *Ideas and Results,* Feferman suggested a very sensible and startlingly obvious objection to the proposal on identity of proofs put forward in the article. It is, in essence, the following: one cannot be expected to accept that a derivation that is more informative than some other in a relevant way is equivalent to it viz. represent the same proof the latter does; and it is obviously the case that, by means of reductions, pieces of information – pieces of what can indeed be regarded as *essential* information – can be simply wiped out of a derivation. The author gives as an example “a derivation D which ends with of $\forall x A(x)$ followed by $(\forall - elimination)$ to get an instance A(t). Let D reduce to D' in n.f. [normal form]. In general, we know more from D than we do from D' and we would say D and D' represent distinct proofs.”

Feferman’s line of thought in this objection is not only far from extravagant, but also apparently intimately connected to a question that arguably helped prompting the interest on the investigation of identity of proofs, which finds good expression in a quote attributed to Kreisel: “What more do we know if we have proved a theorem by restricted means than if we merely know that it is true?” Put roughly, it seems that the initial discussion in the sixties and early seventies on identity of proofs hinges upon the initial assumption that two proofs of a same given result can bring about different pieces of knowledge; hence, it seems quite natural to infer that a proof that clearly yields less knowledge than some other cannot be reasonably considered identical to it – i.e., that what one gets to know by means of a proof is an essential trait of the proof’s identity.

If read in an even more abstract register, Feferman’s objection can actually be seen as a sort of general case instantiated by the objection described in III.2.a.1. The idea is, as stated before, that cutting off or, in general, altering information that might be regarded as essential from a proof is not identity preserving in general; and the variable bonds erased and created by the reductions can be understood as precisely such sort of information.

Notwithstanding the fact that he rejects the normalisation thesis on identity of proofs based on this argument, Feferman does see some significance in the formal relations involved in its formulation. On that matter, he claims:

“Even if it does not settle the relation of identity between proofs, the work described by Prawitz may give simple syntactic explanations of other familiar relations and operations, for example, for the idea of *one proof specializing to another* or of *extracting from a proof just what is needed for its particular conclusion*. Obvious formal candidates for these are the relation: [proof] D followed by $\forall$ -elimination reduces to [proof] D'; and the operation, normalize [proof] D, respectively.”

It is worth noting that Feferman makes this point in a way that might give rise to some confusion. The idea of “extracting from *a proof* just what is needed for its particular conclusion” can be taken in such a way as to demand that one knows beforehand certain things regarding the identity of t*he proof* represented by the derivation under scrutiny. In particular, to justify the claim that normalisation would be a good way of "extracting from a proof just what is needed for its particular conclusion", one would have to know that, say, no part of a normal derivation could be regarded as not "needed"with respect to the particular conclusion of the proof it represents – which would clearly imply that e.g. no two distinct normal derivations could represent the same proof. But such assumptions would not be justified in the absence of a theory on identity of proofs such as e.g. Prawitz's. The possibility of such a reading – which is of course neither the only nor the best possible one – of Feferman's suggestion could be avoided, however, by substituting the notion of "just what is needed" by a notion of *just what is relevant* along the lines described in footnote 22 above.

Be that as it may, Prawitz 1981 p.249 answers to this objection by suggesting that if one sees proofs not "as a collection of sentences but as the result of applying certain operations to obtain a certain end result", his thesis is safeguarded from the threat posed by Feferman's observation – the truth of which he recognises as undeniable, however.

Just what was meant by "the result of applying certain operations to obtain a certain end result" and exactly how this is opposed to "a collection of sentences" is of course not very clear; but this is not a matter on which I intend to dwell right now. Let us rather notice the peculiarity of Prawitz's argumentative move here, which in short consists of stating which informal conception of "proof" is in fact properly described by his theory on identity of proofs.

In the first place, it seems that a more solid interpretation of Feferman's argument is that it is a criticism precisely to the relevance, importance or even acceptability in the context in question of an informal notion of proof the identity of which is insensible to such relevant semantic changes as the one it points at. Hence, Prawitz claim that his formal proposal matches a given informal notion of proof which displays such shortcomings is not really an answer, and does nothing to put his proposal out of the reach of Feferman's point.

Let us nevertheless suppose, for the sake of the argument, that Prawitz interpreted Feferman's criticism in an adequate way, and it really is the case that the reviewer merely failed to understand which informal concept of proof was at stake in his formal proposal on identity of proofs. If this is really the case, then Prawitz could have answered to any criticism whatsoever to his (or indeed any other) formal proposal with smaller or larger amounts of creativity – i.e. by claiming to have intended to describe a certain notion of proof (regardless of how abstruse) that happens to match the formal proposal in question. Regarding this point, Troelstra is quite precise in an observation made in his *Non-extensional equality*:

“If we think of the objects of the theory as ‘direct proofs’ whose canonical descriptions are normal deductions (cf. numbers and numerals in the case of arithmetic), then the inference rules →I, →E correspond to certain operations on direct proofs; the normalisation theorem then establishes that these operations are always defined, and permits us to regard any deduction as a description of a direct proof. The conjecture [i.e. the normalisation thesis] in the direction <= [soundness] then becomes trivially true.

But if the deductions are seen as descriptions of a more general concept of proof (...) (in the same manner as closed terms are descriptions of computations in the case of arithmetic) it may be argued that intensional equality should correspond to (literal) equality of deductions; and then the conjecture is false.

So, without further information about the intended concept of proof, the conjecture in the direction <= [soundness] is meaningless because, as just explained, ambiguous.”

It thus hardly seems that a real debate happened after all. Indeed, it looks like Feferman merely pointed at something that he – quite understandably – deems an essential trait of proofs to which Prawitz’s formal characterisation of identity of proofs does not make justice; and that Prawitz, instead of either somehow disputing the essentiality of this trait to the notion of proof or showing why his notion does not disregard it, was contented to formulate an informal notion of proof that he deems acceptable without further and the identity of which is insensitive to this trait.

Notice that, while voicing the above objection to the normalisation thesis, that significantly shares Feferman’s line of reasoning, Troelstra also argues that the kind of conception of proof Feferman mobilises in his objection suggests us to regard the issue of identity of proofs as somehow reducing to the uninteresting relation of explicit syntactical identity between derivations:

“… if the deductions are seen as descriptions of a more general concept of proof [than that of “direct proof”](…) it may be argued that intensional equality [viz. identity of proofs] should correspond to (literal) equality of deductions; and then the conjecture is false.”

This might indeed be seen as a more explicit expression of the same accusation Prawitz quite subtly suggested about Feferman’s idea on proofs and their identity with the expression “collection of sentences”. Be that as it may, it is at best an overstatement: there is clearly more to be made with Feferman’s point than just abruptly interpreting every syntactical variation as relevant for the semantic value of a derivation *qua* proof. We shall now suggest an adaptation of the normalisation thesis to meet the expectations expressed by Feferman in the mentioned review. As will be clear, the key idea in it is to differentiate between redundant and irrelevant maximal formulas.

a.2.1. (A sketch of) an adaptation of the normalisation thesis to Feferman’s point

The normalisation thesis can be regarded as based upon the understanding that derivations represent proofs the *result* of which is *the semantic value of their end-formulas*, or, more elaborately, a relation of logical consequence between *the semantic value of their end-formulas* and *the semantic value of their set of undischarged top-formulas*. This depiction of the meaning of derivations arguably allows the distortion of a lot of what one could sensibly consider as interesting and informative deductive content encoded in a derivation into nothing but irrelevant detours, which could therefore be eliminated "*salva identitate demonstrationum*" – this is what Feferman's remark in his review of *Ideas and Results* points to. In other words, the normalisation criterion for identity of proofs depends on a rather partial account of what proofs are – most especially, of what proofs prove –, which thus turns it into something as trustworthy and adequate for identifying proofs in certain contexts as a butcher's cleaver is for performing neurosurgery. We should therefore consider alternative ways of understanding OF and FROM what a proof represented by a given derivation is – i.e. what occupies the places of consequent and antecedent in the relation of consequence that is here taken as the result of the proof.

To try and incorporate Feferman’s point in an identity criterion based on normalisation, we start by considering that a derivation could in principle represent a proof of any of its “partial results” - i.e. of any one of the relations of logical consequence between the semantical value of its suitable sets of formulas – that is to say, once we assume that a derivation represents a proof of something from something, we conclude that a derivation is either ambiguous with respect to the proof it represents, or simply represents many proofs the result of which is a relation of consequence between a set of hypotheses and a conclusion. Given the proper restrictions, such a proof could be of anything from anything it contains – and this seems as unbiased, as “general” in spirit as possible in the framework of the basic assumption that the result of a proof is a relation of consequence between assumptions and conclusion. We could also of course just say that the proof represented by a given derivation is the set of all proofs it contains in the former sense.

We can thus start by performing a complete analysis, so to speak, of the deductive information of a given derivation. This can be achieved by writing a list of every subderivation of the derivation under scrutiny – this guarantees that no deductive content shall be lost in eventual reduction processes we decide to perform. After this, we normalise each of these subderivations, in an attempt to separate, as suggested by Feferman, just what was relevant in them for the obtainment of the respective end-formula from the top ones. After that, we can discard all normal forms thereby obtained which are subderivations of some other normal form obtained – for they would consist in superfluous repetitions of deductive information. The set of derivations that comes out of this will clearly contain only the normal forms of subderivations of the original derivation which contain the "necessary material" for the obtainment of each one of the formulas that occur in the original derivation as conclusion of some subderivation, without any repetitions or irrelevancies. Now, this set is to be understood as expressing the identity value of the original derivation. This gives an analysis of the meaning of derivations that preserves their "informational" content as would please Feferman, and quite naturally allows for richer comparisons between them than mere identification or distinction. One could e.g. spot which common normal forms show up in the analyses of two given derivations and understand what they have in common in spite of being eventually distinct. In an analogy, one could explain the difference between Prawitz's approach and the present one in the following way: while the first pictures proofs as paths towards a single end destination that can be either directly or indirectly trod, the second conceives them as a multi-destination paths and analyses them as a "composition", so to speak, of the direct version of all single-destination paths that are trod between its appropriate points. Let us thus try and formulate the idea more clearly by stating the following: *the identity value of a derivation Π is the set of the expanded normal forms Σ'_i of all subderivations Σ_i of Π such that no Σ'_i is a subderivation of some Σ'_j, $i \neq j$.* It is easy to see that the identity value of a derivation is unique, which shows that this way of dealing with identity of proofs yields no ambiguous derivations.

Let us now assume that two derivations represent the same proof iff they have the same identity value according to the definition just given. Let us consider as example of this idea the analyses of the following two derivations Π and Π':

$$\dfrac{\dfrac{\dfrac{[A]^1}{A \supset A}{}_1 \quad A}{A} \quad B}{A \wedge B} \qquad \text{and} \qquad \dfrac{\dfrac{\dfrac{[A]^1 \quad B}{A \wedge B}}{A}}{A \supset A}{}_1$$

In the identity value obtained for the derivation to the left, there will clearly be a derivation which is not in the identity value corresponding to that to the right, and vice-versa: they are, respectively:

$$\dfrac{A \supset A \qquad A}{A} \qquad \text{and} \qquad \dfrac{A \wedge B}{A}$$

All other subderivations, however, are present in both sets. These two derivations are thus not to be identified, but they share most of their deductive information. Now, if we were to consider the following derivation Π'':

$$\dfrac{\dfrac{\dfrac{[A]^1}{A \vee A} \quad [A]^2 \quad [A]^2}{A}{}_2}{A \supset A}{}_1$$

According to the normalisation thesis, Π'' is identical to Π'. Even so, according to the criterion we are now considering, Π'' clearly has less in common with Π' than Π, which e.g. does not even share its end-formula.

In accordance with Feferman's suggestion, then, reducibility assumes here a totally different role: instead of identifying derivations, it has to do with selecting all relevant content in subderivations, so that in the end solely certain innocuous repetitions will be discarded from a derivations identity value. This way of conceiving of identity of proofs thus enables the attribution of an important role to the semantic significance of normalisation in the context of identity of proofs without having to rely on a teleology of proofs as heavy as the one that seems to underpin the normalisation thesis.

b. Claiming the completeness of (αF): the case for (β) or the reasons why Post-completeness (maximality) does not favour the normalisation thesis.

As mentioned previously, point (β) is frequently treated by the literature as a sort of Achilles heel of the normalisation thesis. This much is clear since Prawitz's first formulation of it, which is followed by his explicit acknowledgement that, while “it seems evident” that “a proper reduction does not affect the identity of the proof represented”[26], it is “more difficult to find facts that would support” this half of the “conjecture”[27]. He nevertheless refers us to the text of Kreisel published in the same volume, in which the latter attributes to Barendregt an idea that would be a way out of this situation: a proof of the eventual Post-completeness of the notion of identity determined by the normalisation thesis. The core of the idea is that the Post-completeness of an intended formalisation of the notion of identity of proofs would be a decisive argument for its completeness in case its soundness is granted[28].

To bring the following points to an adequate level of clarity and self-containment, it will be convenient to provide a precise enough explanation of what we are referring to when talking of Post-completeness here. Thus, let ≡ be a relation of equivalence that holds between derivations. We first say that a relation ≡’ is an (Γ, *A)*-extension of ≡ iff there are two derivations Π and Σ of *A* from Γ such that $\Pi \not\equiv \Sigma$ and Π ≡’ Σ. A relation ≡ is understood as Post-complete with respect to (Γ, *A)* iff its only (Γ, *A)*-extension is the trivial relation, i.e. the relation that makes all derivations of *A* from Γ equivalent. A relation ≡ is thus Post-complete in case it is Post-complete with respect to every (Γ, *A).*

The structure of the argument is roughly the following: Suppose that (i) a given relation ≡ is sound with respect to identity of proofs; (ii) ≡ is also Post-complete; and (iii) identity of proofs is not a trivial relation. Now suppose further, for absurdity, that (iv) ≡ is not complete with respect to identity of proofs. Then, by (i) and (iv), there would be more proofs of *A* from Γ identical than those already identified by ≡ – i.e. the set of proofs identified by ≡ would be a proper subset of the set of all identical proofs. But given (ii), this would in turn imply that

26 Some doubt is nevertheless put, as already mentioned, upon the permutative reductions, for they are not based on the inversion principle, which supposedly justifies the purported identity-preserving character of β and η.

27 Regardless of the fact that it has initially been announced as a conjecture, I purposefully eschew the terminology “conjecture” when referring to the normalisation thesis. In contexts that are mathematical enough – such as that of the present discussion of identity of proofs –, a conjecture is something that is, at least in principle, a candidate to be mathematically proved or disproved – or, at the very least, being mathematically shown to be undecidable. This is however obviously not the case of the normalisation thesis, which merely puts forward a mathematically precise *definition* of what being (identical to) a specific proof , say Π, would be, which should be either accepted or rejected (or neither) on ultimately informal grounds.

28 Some have indeed seen the obtainment of such Post-completeness results as the central step in what they call a *proof* of completeness in case of soundness. See the abstract of Widebäck 2001: “The main result of this thesis is that the completeness part of the conjecture is true for the system of minimal implicational logic, provided that soundness can be taken for granted. The result is obtained by first proving that the notion of βη-equivalence is Post-complete. (…)It is then argued that the identity relation on proofs is non-trivial, i.e. that there are non-identical proofs. This proves the completeness part of the conjecture.”

identity of proofs is trivial. By (iii), however, this is absurd. Then we deny (iv) and admit that ≡ is complete. To facilitate understandability, let us organise this argument more graphically:

$$\dfrac{\dfrac{\dfrac{\dfrac{(i) \qquad [(iv)]^{1}}{\textit{The set of proofs identified by } \equiv \textit{ is a proper } \text{subset} \textit{ of the set of all identical proofs.}} \qquad (ii)}{\textit{identity of proofs is trivial.}} \qquad (iii)}{\perp}}{\equiv \textit{ is complete.}}\,{}^{1}$$

Now, supported by this reasoning, and thanks to the already high level of confidence in the soundness part of the normalisation thesis (which amounts to confidence in premiss (i)), the actual obtainment of Post-completeness results concerning the identity relation corresponding to the normalisation thesis (cf. e.g. abstract of Widebäck 2001 and Došen 2003, p.14) (which amounts to an authorisation to affirm premiss (ii)) has been widely regarded as the until then missing stamp of approval on the completeness part of the normalisation thesis; and thus, as a strong reason for accepting the thesis as a whole. There are, however, severe problems with this line of thought, which we shall analyse subsequently. Indeed, we shall argue that the Post-completeness results not only not necessarily favour but also may possibly play against the acceptance of the normalisation thesis, depending on one's assumptions.

b.1) Post-completeness is neither essential nor decisive as an argument for completeness: Firstly, one should remember that the Post-completeness results only yield a "proof" of completeness under the condition that soundness – and also an additional premiss, represented above as (iii) – is assumed beforehand. It is easy to see that, instead of denying premiss (iv) in the last step of the argument above, one could have simply denied (i), i.e. one could have denied *soundness* – a fact which all by itself shows how very mistaken is the idea that such results do *themselves* favour the normalisation thesis. Clearly, one could have reasons to look at the picture upside-down and consider the Post-completeness of the notion of identity determined by the thesis a quite serious drawback of it – say, due to some very clear and appealing example of two non-equivalent derivations which should be considered as expressing the same proof (see e.g. synonymous derivations in section IV.2.a). The condition that soundness is to be "assumed beforehand" has, thus, a very strong meaning: far more than merely assuming the soundness part of the thesis as one of the initial hypotheses that could in principle be somehow discharged, one must commit to sustaining it throughout the argument so that the desired conclusion, i.e. completeness, can be reached. In this argument, completeness *depends* on soundness, regardless of how reasonable it is to assume soundness.

Now, one may accept soundness as a premiss for different reasons (the seemingly strongest of them, the proof-theoretic semantic framework of Prawitz and Dummett, already described in section III.2.a, is of course highly questionable; but this fact is now off the point). One may also do this much for no reason at all, only for the sake of the argument. Each of these cases will be treated respectively in the next two subsections.

b.1.1) Let us then suppose that our acceptance of soundness stems from some very solidly grounded motivation – say, the semantical background considerations presented in section III.2.a. These suggest that reductions should be deemed identity preserving due to their being based on the so-called inversion principle. But if that is so, there are no other derivations to be considered as reasonable candidates to being identical other than those already identified by the equivalence relation determined by the reductions involved in normalisation: for it is clear from the very outset that no other two derivations may differ only up to at most what the inversion principle determines as "detours". So, in this case, the reason to affirm soundness already grants us completeness – for an adequate formulation of the inversion principle should allow us to systematically arrive at the complete set of accepted reductions.

To provide ourselves a more palpable illustration of this argument, let us remember how Prawitz formulates the inversion principle: "*Let α be an application of an elimination rule that has B as consequence. Then, deductions that satisfy the sufficient condition (…) for deriving the major premiss of α, when combined with deductions of the minor premisses of α (if any), already "contain" a deduction of B; the deduction of B is thus obtainable directly from the given deductions without the addition of α."*

As already pointed out in section III.2.a, the principle refers first to the following deductive pattern:

$$\dfrac{\dfrac{\Pi_1, \dots, \Pi_n}{*^n(A_1, \dots, A_{n_1})}{\scriptstyle *^n\text{-}introduction} \qquad \begin{matrix}\Gamma_1 \\ \vdots \\ C_1\end{matrix} \;\; \cdots \;\; \begin{matrix}\Gamma_m \\ \vdots \\ C_m\end{matrix}}{B}\,\alpha$$

Its claim is roughly that B, the conclusion, could always be obtained by eschewing the introduction of $*^n(A_1, \dots, A_n)$ in the following way:

$$\begin{array}{c} \Gamma_i \\ \vdots \\ C_i \\ \vdots \\ \Pi_j \\ \vdots \\ B \end{array}$$

As it happens, the inversion principle does evidently not hold for just any constant we could make up by bundling together randomly picked introduction and elimination rules. The result of substituting an instance of the first deductive pattern for one of the second for a given constant * could well turn out not to be a derivation. But it is quite clear that, for each of the intended constants, it is possible to substitute the respective instances of the first deductive pattern for one of the second while preserving the "good formation" (as well as the rest of the structure) of the derivation where they occur untouched. In other words, the reduction steps associated with each constant are direct instances of the more abstract reduction scheme we call the inversion principle. But that is not all. It is also clearly the case that other possible reductions concerning the intended constants are not direct instances of the inversion principle.

Of course, one could question whether or not the inversion principle is the only principle capable of motivating identity preserving transformations on derivations –- just as e.g. Kant believed that *dictum de omni* and *dictum de nullo* were the only principles ultimately capable of justifying the logical validity of inferences of reason. It is a matter of strong controversy whether or not some alternative reductions already suggested in the literature do qualify as identity preserving in some sense[29]. For instance, it could be argued that substituting the application of the introduction rule for an application of *ex falso quodlibet* with the same conclusion in the usual redexes yields yet another kind of redex, to be reduced without loss of identity in the same way as the correspondent usual ones. It could also be argued that synonymous derivations in the sense of section IV.2.a are identical. In any case, for all such possibilities, the same kind of remark made with respect to the inversion principle applies: they are ultimately the expression of principles that can – at least in principle –- be expressed in such a way that we know quite precisely whether or not there could be more

29 This qualification – "identity preserving in some sense" – is actually of vital importance to the matter under discussion. The Post-completeness of a given set of principles deemed identity preserving in one sense means nothing with respect to its compatibility with another one deemed identity preserving in some other sense. Indeed, there should be no question regarding their compatibility: it would be comparing apples and oranges. All considerations in this subsection start from the assumption that all potentially identity preserving principles are so in the same sense, even if for different reasons.

ways to identify derivations by virtue of them apart from the ones at hand. Thus, with each principle that motivates the identification of distinct derivations, the reason to support its soundness should already bring with itself the argument to support its completeness.

Those remarks considered, there are however two points that still make Post-completeness look appealing in this scenario. The first is the one that has directly to do with completeness: it is the, say, sceptical consideration that there may be more principles fit to identify distinct derivations than the ones we (can) understand as such. But in such a case, we would be in a situation where there is no reason at all at hand to grant soundness to any of those supposed principles. The Post-completeness of the set of justified principles could thus serve as means to preclude the possibility that derivations no one has ever thought about as even possible candidates to being identical are understood as identical. Giving Post-completeness as an argument to support completeness in such a setting is thus an inversion of the burden of proof, so to speak: for one should first come up with a reason that may satisfactorily justify the identification of further derivations than those already identified by justifiedly accepted principles before meaningfully challenging the completeness of such principles with respect to identity of proofs. The second reason concerns completeness only indirectly; it has to do with the possibility of "coexistence", so to speak, of a given set of principles of identification of derivations with others. Let us assume we have two sets of principles that may induce equivalence relations between derivations. It may well be the case that the eventual union of these principles provokes a collapse; namely, that the equivalence relation induced by their union turns out to be trivial, even in case this does not happen when these principles are taken separately. Suppose now that we have good reasons to understand both of them as being among those deemed identity preserving, and also that their union causes the kind of collapse described. Then, clearly, we must sacrifice something to avoid contradicting ourselves: either we give up the soundness of at least one of these sets of principles, in spite of having equally good reasons to understand them as sound; or we give up the idea that identity of proofs is not trivial. Now, the eventual Post-completeness of a given such set of principles poses, before anything else, a particular, in a sense limiting case of this kind of short blanket situation. Namely this: the union of a Post-complete set of principles with any other will cause the collapse of the corresponding equivalence relation between derivations. So, if a Post-complete set of principles is deemed sound with respect to identity of proofs together with any other further set of principles – even if for equally good and strong reasons –, then the choice must be made: either, on the one hand, we discard its soundness or the soundness of all further sets of principles; or, on the other, we give up the non-triviality of identity of proofs. So, as claimed above, the eventual Post-completeness of a given set of principles indeed gives us information on how fit it is to coexist with others in the set of those which one deems identity preserving, namely: not at all. Whether this should count as a reason to take it as complete and disregard the reasons why others should also be deemed sound with respect to identity of proofs; or as a reason to deny its soundness with respect to identity of proofs in spite of the reasons we have to sustain it; is a matter for which its Post-completeness is all by itself utterly irrelevant.

b.1.2) These last remarks take us already to the case in which we decide to grant soundness to a given set of principles with respect to identity of proofs for no systematic reason, e.g. because the reductions seem intuitively identity preserving, or even arbitrarily, say, just for the sake of argumentation. In such a setting, an argument for completeness is really due. Again in this case, though, Post-completeness would not suffice as a reason to accept the thesis as a whole. For one could in principle have other sets of reductions resp. equivalent derivations to which one could be willing to grant soundness for equal reasons, and which one could simply decide to favour as sound while discarding the assumption of the soundness of the previous, Post-complete one. In principle, such sets could even be proved to be also Post-complete; which would then preclude even the possibility of fallaciously using Post-completeness as a reason *ad hoc* to favour the election of any of the alternatives. To facilitate understandability, let us try to describe this reasoning by structuring the premisses of the argument as made just above, and then display it more graphically.

Thus, just as in the case of ≡, let us start by supposing that (i’) a given relation ≡’ is sound with respect to identity of proofs. Once more analogously, suppose also that (ii’) ≡’ is Post-complete; and, as before, that (iii) identity of proofs is not a trivial relation. Suppose further that (iv’) ≡’ is not complete with respect to identity of proofs. Obviously, as in the case of ≡, by (i’) and (iv’), there would be more proofs of *A* from Γ identical than those already identified by ≡’ – i.e. the set of proofs identified by ≡’ would be a proper subset of the set of all identical proofs. But given (ii’), this would in turn imply that identity of proofs is trivial. By (iii), however, this is absurd. Then we deny (iv’) and admit that ≡’ is complete. Graphically:

$$\dfrac{\dfrac{\dfrac{\dfrac{(i') \qquad [(iv')]^1}{\textit{The set of proofs identified by } \equiv' \textit{ is a proper } \text{subset} \textit{ of the set of all identical proofs.}} \quad (ii')}{\textit{identity of proofs is trivial.}} \quad (iii)}{\bot}}{\equiv' \textit{ is complete.}}\,1$$

Now, let us call these two arguments – the one that concludes the completeness of ≡ and the one that concludes the completeness of ≡' – Π and Π', respectively. Let us also note explicitly the assumption that defines the role of ≡' in the argument, namely that (i≠) ≡' and ≡ do not have the same extension. Now, putting Π and Π' side by side together with (i≠) clearly leads to absurdity. We thus face the following situation: to avoid absurdity, we should deny (i), (ii), (iii), (i'), (ii') or (i≠). But none of the options seems justifiable: (ii) and (ii') are treated as facts; (iii) is what gives the whole discussion its *raison d'être*; (i≠) is merely devised to prevent the identification of ≡ with ≡'; and there is no reason to favour (i) instead of (i') or vice-versa. Not even the possibility of seeing in the mere Post-completeness of ≡ some kind of motivation for denying (i') or (i≠) *ad hoc* – which in any case would be completely insufficient as an argument – is at hand, for the hypothesis is exactly that ≡' is also Post-complete (and this also holds vice-versa, of course). The choice of any of the two alternatives in such a setting would then be unavoidably arbitrary. Graphically:

$$\dfrac{\dfrac{\begin{matrix}\{[(i)],[(ii)],[(iii)]\}\\ \Pi \\ \equiv \textit{ is complete}\end{matrix} \qquad \begin{matrix}\{[(i')],[(ii')],[(iii)]\}\\ \Pi' \\ \equiv' \textit{ is complete}\end{matrix} \qquad [i \neq]}{\bot}}{???}$$

Thus, in the first-place, while (ii') remains a possibility, the burden is upon the shoulders of the proponent of the Post-completeness argument to show that it can be safely denied – i.e. to avoid the possibility that her argument leads to a dead-end absurdity, the proponent must still show that no other relation of equivalence between derivations that can be considered sound with respect to identity of proofs is Post-complete. Secondly, even if she manages to ascertain this much, the reward is too poor: a certificate that no other equivalence relation can be considered complete for the same *ad hoc* and – no matter how mathematically convenient – philosophically insatisfactory reason that the one proved Post-complete can.

The exposition just carried out portraits far more faithfully the significance of Post-completeness results concerning this discussion than the usual device of using them as pieces of propaganda in favour of the normalisation thesis or some other formal characterisation of identity of proofs that enjoy this mathematical property.

b.2) The conclusion drawn from premisses (i) and (iv) that there would be more proofs *of A from Γ* identical than ≡ identifies is only possible in the presence of an additional, implicit assumption: that identity of proofs is understood, from the very outset and independently of the normalisation thesis itself, as a relation that identifies only proofs carried out by means of derivations which share the same conclusion and undischarged hypotheses. This actually amounts to a way of placing the normalisation thesis in the category "restricted to the result" of our taxonomy. Let us now briefly consider the significance of this taxonomic decision.

The formalisation of the core notion of irrelevance discussed above as the famous "detours" of Prawitz is *per se* of an ambiguous nature. On the one hand, one might regard this formalisation as somehow conceived within the limits of a previously accepted restriction upon the notion of identity of proofs which places it under the category *restricted to the result* in our taxonomy – a restriction that would not only be understood as independent of how the notion of irrelevance should be specifically understood in the context of identity of proofs, but also as having the very notion of irrelevance under its jurisdiction. On the other hand, one might as well see this picture turned upside-down, and say that this formal rendering of the notion of irrelevance is not previously restricted by any means, and that the restriction to same sets of end- and undischarged top-formulas imposed on the notion of identity of proofs that results from its adoption is indeed a *consequence* of how irrelevance alone should be understood. In this second case, it is fair to understand the normalisation thesis as belonging to the *unrestricted* category in the taxonomy.

The impression that this is a somehow innocuous, negligible conceptual distinction is quickly corrected when one considers e.g. precisely the argumentative strategy of appealing to Post-completeness to advocate the completeness of the notion of identity yielded by the normalisation thesis. If one supposes that the second of the options given above is the case, for instance, then the Post-completeness of the formal relation candidate to the role of counterpart of identity of proofs is completely irrelevant: for in such a case, there would be no reason to determine beforehand – i.e. it could not stem from an intensional restriction upon the notion of identity of proofs itself – that proofs carried out by means of derivations which do not share both end-formula and assumptions are distinct. Given that the Post-completeness results involved in the argument favouring completeness concern solely extensions of the notion of identity yielded by the normalisation thesis which identify more proofs *of given A from given Γ* than the notion of identity itself does, they become obviously insufficient as an argument to support the completeness of a notion that may in principle identify proofs which do not share both end-formula and assumptions.

This shows that the argumentative strategy under scrutiny depends crucially, as already claimed, on the supposition of the first hypothesis – indeed, of a very strict form of it, namely, that identity of proofs identifies only proofs carried out by means of derivations which share conclusions and undischarged hypotheses. Given some very basic issues already discussed along this work, it is not only not necessary but also rather strange – even though very usual – to impose this sort of restriction upon the notion of identity of proofs right from the outset of its investigation. Hence, in the current absence of good arguments to sustain such a move, it has nothing but the idea of identity of proofs outlined by the normalisation thesis itself to motivate it. In such a case, thus, the Post-completeness argument involves circular reasoning, bearing no probative import whatsoever with respect to the completeness of the normalisation thesis. Otherwise, the burden lies upon the shoulders of those willing to maintain the Post-completeness argument to provide foundations to the specific way of taxonimising identity of proofs upon which it depends.

These observations should be enough to make it clear that the normalisation thesis, very similarly to e.g. the Church-Turing thesis, is not to be regarded as a mathematical conjecture – for no matter how mathematically precise its formulation is, it is, at least as far as *mathematics* is concerned, of a definitional rather than of a descriptive nature. The point made in item b.1) shows that the Post-completeness results concerning the normalisation thesis have been discussed in a severely partial way, and that they are not to be counted as arguments in favour of the normalisation thesis. The point made in item b.2) in turn shows that taxonomic matters play a central role in arguing for or against a thesis on identity of proofs, which also undermines hope for any attempts to *prov*e or *disprove* them, even if only partially or conditionally, in a way that is simultaneously mathematical and significant enough.

c. Proofs without irrelevancies: (γ), (strong) normalisation and uniqueness of normal form

Premiss (γ) is not, strictly speaking, a constituent of the normalisation thesis itself. This becomes clear when we consider its official formulation: “Two derivations represent the same proof iff they are (βη-)equivalent”. There is no mention whatsoever to normalisation or normal forms – and indeed, given the arguments considered so far, it seems we would not have any (non *ad hoc*) reasons to regard β- or η- conversions as any less identity preserving in case normalisation failed for them. Widebäck (2001, p.76), for instance, seems to acknowledge the secondary role of the obtainment of normalisation to the specific matter of identity of proofs rather in this spirit when he states, in his discussion of Ekman’s reduction, that “from our point of view, the failure of weak normalization is, at most, a technical inconvenience”.

The idea conveyed by this premiss is nevertheless quite ironically the reason why we refer to it as the *normalisation* thesis in the first place. The fact is that normalisation *does* hold for Prawitz’s reductions, which means that premiss (γ) works more as a corollary than as a proper addition to the normalisation thesis – and this is really one of its most remarkable features. The informal, motivational counterpart of the normalisation thesis does not entail premiss (γ) at all; but in rendering premiss (α) as (αF), the normalisation thesis outlines a notion of irrelevance such that it is necessarily the case that all proofs can be mechanically freed from their eventual purported irrelevancies.

But the acceptance of (γ) is not all by itself enough to turn the idea behind normalisation thesis into something that can be formulated as suggested in the beginning, namely, the claim that any two proofs are significantly different if and only if they are different *even when given without any irrelevant steps/ features*. For instance, even assuming that a proof can be freed from all its irrelevant features, this does not mean that this can be done in only one way, nor that all ways of doing it would have the same outcome. So, it could in principle well be the case that a single derivation might end up split in two different ones when freed from its irrelevancies in two different ways. But this possibility seems at odds with the idea in the formulation above: for there could then be proofs different from themselves.

This is where the uniqueness of normal form shows its importance. There are no two different ways of completely freeing a derivation from its irrelevancies such that they split it into two different irrelevancy-free versions of itself; every reduction sequence of a derivation that terminates does so leading to the very same normal form in the end. The fact that strong normalisation holds, i.e. that every sequence of reduction terminates, makes the picture even more seductive: for there is then, additionally, no way of freeing a derivation of its irrelevant features that may not lead to an irrelevancy-free version of itself. To sum up, every possible way of freeing a derivation from its irrelevant features – which is by necessity built exclusively out of identity-preserving transformations – leads to a same and unique irrelevancy-free version of itself. This version – its normal form – is thus given the role of a canonical "representative" of the identity values of the derivations which reduce to it, which are thus of course, <u>given the assumption of the non-ambiguity of the identity-value of a derivation</u>, deemed to be the same.

But as innocent as this assumption may seem, it really stands upon thin ice in this discussion. Let us, for instance, consider the following derivation:

$$\dfrac{\dfrac{\perp}{A \vee B} \qquad \begin{matrix}[A] \\ \vdots \\ C\end{matrix} \qquad \begin{matrix}[B] \\ \vdots \\ C\end{matrix}}{C}$$

Since it introduces a complex formula and eliminates it immediately after, one might look upon it as displaying a structure that fits the idea of a "detour" involved in the normalisation thesis (Indeed, the occurrence of $A \vee B$ is a maximum formula according to the definition given by Prawitz 1965, p.34). There would be, however, two obvious ways in which we could get rid of this:

$$\begin{matrix}\dfrac{\perp}{[A]} \\ \vdots \\ C\end{matrix} \quad \text{or} \quad \begin{matrix}\dfrac{\perp}{[B]} \\ \vdots \\ C\end{matrix}$$

Letting ourselves be guided by the motivating ideas of the normalisation thesis, it seems quite sensible to see these two simplified derivations as significantly different from each other – at the very least, in any case, their normalisation sequences clearly do not converge. This failure of uniqueness of normal form could be taken to point at a case of *ambiguity* of a derivation with respect to its identity value. This sort of phenomenon does not necessarily require the presence of "detours"; indeed, it is $\perp$ and $\vee$ that seem to be the main ingredients in these cases. For instance, consider the derivation:

$$\frac{\perp}{A \vee B}$$

Motivated by the ideal of separation of the roles of the logical constants and the simplicity of inference rules (as e.g. described by Prawitz in section II.2.1 of his *Ideas and Results* 1971 paper), one could regard this derivation as further analysable, so that each of its steps shall be atomic and each constant shall be introduced by its correspondent, "meaning-giving" rule. Thus, there would be again two – and again arguably significantly different – ways of expanding it in order to obtain such a goal:

$$\frac{\dfrac{\perp}{A}}{A \vee B} \quad \text{and} \quad \frac{\dfrac{\perp}{B}}{A \vee B}$$

This time, however, the divergent solutions – that similarly point at what can be regarded as an ambiguity – do not come about by virtue of any "detours"; all derivations involved are, in fact, normal.

For the cases considered, one could always claim, in the spirit of some category-theoretic considerations on the matter, that there should be no distinction between derivations of a given formula from absurdity, given that the latter is to be considered an *initial* object. Kosta Došen, for instance, calls the denial of the initiality of absurdity "a desperate measure, not in tune with the other intuitions underlying the normalization conjecture" in his 2003 *Identity of Proofs based on nomalization and generality*, p.19 – but this might well be regarded as, at least, an overstatement. Another way of making the problem disappear is to restrict from the outset the application of absurdity rule to atomic conclusions only (see e.g. Prawitz 1971, pp. 242, 248) – but this seems, again, an *ad hoc* device, hardly in tune with the *ex falso quodlibet* understanding of absurdity. To advocate for the separation of the roles of logical constants, one should be expected to be able to *show* that absurdity can be restricted to atomic cases only; but this would rehabilitate the examples above involving disjunction.

3. A notion of propositional identity based on that of identity of proofs proposed by Kosta Došen

I believe this to be a particularly good opportunity to make some comments with respect to a notion of "propositional identity" put forward by Kosta Došen as seen in connection with the doctrine on identity of proofs under scrutiny. Došen's general idea is that one can solve the issue of when two sentences are synonymous viz. represent the same proposition once a criterion for identity of proofs is established; more precisely, two sentences A and B should be taken as synonymous if and only if they are *isomorphic*, i.e. there is a composition of a proof Π of B from A with a proof Π' of A from B that is identical to the identity proof of A from A, and a composition of Π' with Π that is identical to the identity proof of B from B (where "identity proof" means a proof such that if composed with any other proof, either as assumption or as conclusion, yields a proof identical to the original one). The intuition behind the idea is briefly described by the author in *Identity of Proofs based on Normalization and Generality* in the following fashion:

"That two sentences are isomorphic means that they behave exactly in the same manner in proofs: by composing, we can always extend proofs involving one of them, either as assumption or as conclusion, to proofs involving the other, so that nothing is lost, nor gained. There is always a way back. By composing further with the inverses, we return to the original proofs."

To further illustrate the notion of synonymy outlined by his formal definition, let us consider the two examples the author himself mentions. First, let us take $A \wedge B$ and $B \wedge A$. It is easy to see that they are isomorphic:

$$\dfrac{\dfrac{\dfrac{\dfrac{A\wedge B}{B} \quad \dfrac{A\wedge B}{A}}{B\wedge A}}{A} \qquad \dfrac{\dfrac{\dfrac{A\wedge B}{B} \quad \dfrac{A\wedge B}{A}}{B\wedge A}}{B}}{A\wedge B}$$

This derivation clearly reduces (and is therefore identical) to the identity proof of $A \wedge B$ from $A \wedge B$ (namely $A \wedge B$). One can also clearly build an exactly analogous derivation that reduces to the identity proof of $B \wedge A$ from $B \wedge A$ by simultaneously substituting the occurrences of B for A and vice-versa. This means that $A \wedge B$ and $B \wedge A$ are isomorphic. Now, the same could not be said about *A* and $A \wedge A$:

$$\dfrac{\dfrac{A \wedge A}{A} \qquad \dfrac{A \wedge A}{A}}{A \wedge A}$$

As Došen observes, in the two occurrences of the proof Π of *A* from $A \wedge A$ in the composition above, the projection used, no matter which, must have been the same; for they are, I repeat, *copies* of *only one* given proof Π of *A* from $A \wedge A$. This means that the composition of Π with the proof of $A \wedge A$ from *A* yields no redex, and therefore does not indeed reduce to the identity proof of $A \wedge A$ from $A \wedge A$. Graphically this can be seen even more clearly:

$$\begin{array}{ccccc} A \wedge A & & A \wedge A \\ \ddots & & \ddots \\ A & & A \\ \ddots & & \iddots \\ & A \wedge A & \end{array} \quad \text{or} \quad \begin{array}{ccccc} A \wedge A & & A \wedge A \\ \iddots & & \iddots \\ A & & A \\ \ddots & & \iddots \\ & A \wedge A & \end{array}$$

In words: the two occurrences of *A* in the end-formula are connected to the same occurrence of *A*, respectively the left and the right one, in the copies of the top-formula in the graphs above. To use yet another intuition: in the first case, the proof generalises to a proof of *A* from $A \wedge B$, and in the second, to a proof of *B* from $A \wedge B$. To represent a proof that would reduce to the identity proof in this case, the graph would have to connect each occurrence of in the end-formula, the left and the right one, with its correspondent occurrence of in the top-formulas:

$$\begin{array}{ccc} A \wedge A & & A \wedge A \\ \ddots & & \iddots \\ A & & A \\ \ddots & & \iddots \\ & A \wedge A & \end{array}$$

In such a case, though, it would represent the composition not of one, but of two distinct proofs of *A* from $A \wedge A$ (namely the ones obtained by the two distinct projections) with the proof of $A \wedge A$ from *A*.

Now, this quite novel way of approaching synonymy of propositions envisaged by Došen is particularly sensitive to the same taxonomic issues that affect the Post-completeness argument. Suppose, for instance, that we take normalisation thesis to be restricted to the result, and that we accept the corresponding notion of identity of proofs.

Then, given the remarks made in the previous sections (especially III.2.a.1.1) , we already are comitted to a specific syntactical interpretation of the notion of result of a proof, namely: that the result of a proof carried out by a given derivation is the "intension" that corresponds exclusively to its end formula. Were it otherwise, there would then be ways to identify proofs expressed by derivations that do not share an end formula. In any case, if that will not be granted, it should at least be granted that either this or one of the interpretations based on the observations regarding derivations's principal types is implied by the normalisation thesis. Now, if we additionally accept Došen's idea on synonymy – in his words, "identity" of propositions, which is rather more specific, as our discussion concerning synonymy ahead shall indicate – then e.g. $A \wedge B$ and $B \wedge A$ are, as exemplified above, identical as propositions. This arguably means that the intension that counts as result of a proof of . $A \wedge B$ is different from the proposition expressed by . $A \wedge B$ – viz. more specific than this proposition. The picture is of course coherent, even though one might in this context find it unusual that the result of the proof carried out by a derivation cannot be equated to the proposition expressed by its end formula. It makes one wonder what conception of result – since it is not that of proposition – is being aimed at after all.

In our taxonomy, it would fall under the propositional category, in the section "intension stricter than a proposition"; so, one could consider something as e.g. a specific formulation of a proposition. But of course, one could always just answer "whatever it is that one is willing to call the intensional content exclusively expressed by a formula". But rather than a dull answer, this suggests that Došen's idea could indeed be turned into an alternative to Carnap-style conceptions of intensional content. With the proper handling of taxonomy related issues, it could be taken to define the very notion of result of a proof as the strictest intensional content expressed by a formula – and this would have as a mere consequence that there is a one to one correspondence between formulas and such intensional contents. Based on that, one may proceed to understand synonymy viz. identity of propositions – as well as of course, the concept of proposition itself – as a strictly intensionally defined notion, which corresponds precisely to the second strictest interesting sense in which we should take formulas as sharing an intension.

IV. Normalisation thesis, identity of proofs and synonymy

In this section, we will try to approach normalisation thesis as an answer to the question regarding identity of proofs from the perspective of *synonymy* – i.e., by taking the aimed relation of "identity" between proofs to be indeed a particular case of synonymy. This approach is particularly interesting, since it seems perfectly in tune with what is probably the most traditional way of conceiving of proofs, namely, as *linguistic* expressions. This will of course demand some brief independent reflection on the notion of synonymy itself. Among other things, this part of the study shall allow for interesting comparisons between the normalisation thesis and other possible approaches to identity of proofs motivated by different accounts and understandings of synonymy.

1. A brief conceptual historicisation of the notion of synonymy

It is a matter of uncertainty whether it was Aristotle or Speusippus the first between these two ancient thinkers to write on synonymy. In any case, the fact that both did so and that it is probable that the first influenced the treatment of the second significantly is enough to show how long actual discussion on the matter has been taking place within the literature. For no reason other than my complete lack of knowledge of Speusippus' works, I shall take Aristotle's developments as a starting point. Coincidentally enough, by the way, the explanation of the "–onymies" (viz. homonymy, synonymy and paronymy) is also the starting point of Aristotles treatise *Categories*. I now quote his explanation of synonymy *in extenso*, as translated by J.L. Ackrill:

"When things have the name in common and the definition of being which corresponds to the name is the same, they are called *synonymous*. Thus, for example, both a man and an ox are animals. Each of these is called by a common name, 'animal', and the definition of being is also the same; for if one is to give the definition of each – what being an animal is for each of them – one will give the same definition."[30]

30 J.L. Ackrill, in *Aristotle's Categories and De Interpretatione*, (Oxford, 1962).

The first period of this passage – where the proper definition of synonymy is given – is rendered into English differently by J. Barnes[31]:

“Items are called synonymous if their name is common and in addition the account of their essence in respect of the name is the same.”

After giving a more precise formulation to Aristotle’s definition of synonymy, Barnes adds an explanatory note, apparently with the intention of avoiding what he sees as a misunderstanding favoured by many commentators of Aristotle:

“According to the commentators, Aristotelian homonymy and synonymy are properties of 'things' rather than of words; but that is false – or at any rate, it is misleading. What is true is this: if the members of a group of items are homonymous or synonymous (...), it ‘does not follow from that fact that they are all words (nor, more generally, that they are linguistic items of some sort or other). But neither does it follow that none of them is a word or a linguistic item of some sort.”

Indeed, Ackrill not only makes explicit use of the word ‘things’ in his English version of Aristotle’s definition of the “–onymies”, but also says in the notes that follow his 1962 translation of the *Categories* that:

“The terms ‘homonymous’ and ‘synonymous’, as defined by Aristotle (…), apply not to words but to things. Roughly, two things are (…) synonymous if the same name applies to both in the same sense.”

I believe there is good reason for sustaining both positions; and this is due precisely to the fact that they are not conflicting at all, rather unlike what it might seem at first sight. Untying this little interpretive knot will help us understanding better the notion of synonymy put forward by the Philosopher. A brief look at the Greek version[32] of the variously translated definition of synonymy shall help us a great deal in performing this task:

“συνώνυμα δὲ λέγεται ὧν τό τε ὄνομα κοινὸν καὶ ὁ κατὰ τοὔνομα λόγος τῆς οὐσίας ὁ αὐτός”

There is, as one can see, no noun in the Greek version equivalent to the words “items” in Barnes’ or “things” in Ackrill’s translation. These words are used to render a construction that finds no faithful parallel in English, based in the use of the plural relative pronoun *ὧν*. Roughly, this pronoun works in the formulation as a mere “position-holder” for that to which the relation of synonymy applies. In a forceful attempt to show this working in “English”, it

31 See Barnes, J., 2012, *Speusippus and Aristotle on homonymy*, in Logical Matters: Essays in Ancient Philosophy II, pp.284 – 311, Oxford University Press, Oxford
32 Extracted from L. Minio-Paluello, 1949 Oxford.

would look more or less like this: “Synonymous are called those of which the name is common” – which almost automatically leads one to either assume that the range of the pronoun is groups of two or more people, or to ask: “those what?”. The first alternative is obviously inappropriate; the question in the second one, however, can be adequately answered with any term that is general enough to stand for whatever Aristotle may have referred to with his formulation. ‘Items’ or ‘things’, in fact, seem fit to play such a role very well. We could thus work with the following version: “Synonymous are called those things of which the name is common and of which the account of the essence according to the name is the same.”

Now, what Barnes denounces as misleading about the commentators remarks on Aristotle’s notions of the “-onymies” seems to be intimately connected to the old philosophical vice of attaching some special sort of ontological import to words such as ‘thing’ and ‘object’ that prevents them to refer to things (!) such as words and “linguistic items” – or at least make them refer preferentially to things other than those. It is of course needless to say that the last sentence performatively challenges such an account. It would also be unfair if Barnes credited any misunderstanding of this simple issue to Ackrill himself – which is one of the commentators he addresses, as he explicitly mentions in a footnote. In the same note Barnes refers to in his text, Ackrill, regarding precisely this issue, warns that “It will be necessary in the translation and notes to use the word ‘things’ as a blanket-term for items in any category. It often represents the neuter plural of a Greek article, pronoun, &c.”; thus both explaining why a noun shows up in the translation which is absent in the Greek version and showing explicitly that its range is not intended to be limited to some particular set of categories.

Still, why is it that Ackrill and so many others stress that the proper understanding of Aristotelian synonymy is that it applies “not to words but to things”? Well, a good reason for such an attitude is that, although Aristotle’s definition allows synonymy to apply to things in general – which evidently encompass words and linguistic items, as Barnes stresses –, it will never apply to words *as words*, but rather *as things*. What is meant by this clumsy phrase is that words can be synonymous in Aristotle’s sense by virtue of the fact that they are *things that can somehow be the meaning of linguistic items*, and not because they are themselves provided with some meaning – i.e., exactly the opposite of the common contemporary use of the notion of synonymy, according to which it is understood as an essentially semantical relation – i.e. it only applies to things *inasmuch as they have meaning* – that, according to our instrumental and provisory explication, holds whenever the meanings of the things considered is in some sense equivalent. So, what is characterised by the insistently repeated line “things, not words” is an *intensional* rather than an extensional distinction between the Aristotelian and the common contemporary understandings of synonymy. Indeed, from an extensional point of view, both notions can be taken to apply to both “things” and “words”: if I consider e.g. a hammer, it is Aristotle-synonymous with a screwdriver with respect to the *ὄνομα* ‘tool’. For a quite different reason, the hammer and the screwdriver may be considered synonymous in the usual sense because both may be seen as devoid of meaning. This very same hammer may, on the other hand, be deemed synonymous in the very same usual contemporary sense with a written occurrence of the letter ‘T’ (in case, say, it is creatively put before two gears and a yardstick folded to form a right angle to form the word ‘tool’). A written occurrence of the letter ‘T’, in turn, may, because of the very same iconic resemblance with the tool, be deemed synonymous with the word “hammer”. In short: if one looks extensionally upon this matter, there are in principle no boundaries beyond which creativity may not lead. Being a linguistic or a non-linguistic item is an essentially intensional feature, which anything – i.e. any extension – may or may not in principle display. Barnes’ warning is thus less prone to generate – instead of avoiding, as he certainly intended – even more misunderstandings when formulated in these terms: the observation that Aristotle’s synonymy applies to things and not to words should by no means be taken extensionally, but rather intensionally. When one considers the points just exposed, this seem to be indeed quite obvious.

The important intensional distinction between Aristotle’s notion of synonymy and ours should not however be interpreted as proof that they have nothing to do with each other. In the first place, it should be observed that contemporary use of synonymy is at least extensionally equivalent with a particular case of Aristotle’s concept as defined in the Categories. This is indeed quite easy to see, and can be explained as follows: given that a thing T has a certain meaning, take that it is possible to build a predicate (i.e. an *ὄνομα*) O that is common to all and only the things that are equivalent to T in meaning. In such a case, one can say that something is synonymous to T in the usual sense if and only if it is Aristotle-synonymous to T with respect to O. The idea is quite trivial: if the relation of semantical equivalence may be “simulated” – and therefore indirectly attributed – by means of a predication, then it can be defined as a particular case of Aristotelian synonymy. Secondly, one should not overlook the fact that one of the most visible features that distinguish the two conceptions of synonymy – namely, the fact that ours is an n-ary equivalence relation that may be applied to n things while Aristotle’s is a binary relation between a set of things and an *ὄνομα* – is not but a merely apparent difference; it concerns the formulations rather than the notions themselves. This becomes clear once one observes that the role played by the *ὄνομα* in Aristotle's synonymy is that of a *criterion* according to which the things in the set considered are or are not equivalent. Of course, usual attribution of synonymy is also in accordance to *some* criterion – although this is not commonly made explicit as in Aristotle, which contributes to the frequent misunderstandings and disagreements on the matter. Indeed, it is as if contemporary debate on synonymy somehow always presupposes that there is one correct criterion of equivalence – the nature of which is still unrevealed. More shall be said on this specific point later (section 4.2).

There is nevertheless still a third remark regarding the relation between these two notions of synonymy which makes them look even more similar to one another, namely: none of the things to which Aristotle's notion of synonymy may apply was seen by the philsosopher as "devoid of meaning". Just as our linguistic expressions in general express something – senses, if you will –, so can things in general be looked upon as "expressions" of their respective *essences* within Aristotle's conceptual economy. What I claim – and again neither could nor shall defend here – is that there is an intrinsic *transitivity*, which contemporarily is usually attributed only to the linguistic realm (as in e.g. what *means* vs. what *is meant*), that pervades, and, what is even stronger, characterises the structure of the ontological realm according to the Aristotelian doctrine (as in, e.g. *appearance* vs. *essence*).

Just to give the reader a brief indication of what this claim is based upon, let us consider the Greek expression occurring in the mentioned definition of synonymy in the *Categories* translated by Ackrill as "definition of being" and by Barnes as "account of essence", namely, "*λόγος τῆς οὐσίας*". One is to take seriously the observation made by Ackrill that to understand Aristotle's notion of synonymous things as things to which the same name applies in the same sense is rough. This shortcut to a contemporarily more accessible rendering of the Stagirite's notion indeed involves an unjustifiable anachronism, which consists in implicitly identifying *λόγος τῆς οὐσίας*, i.e., definition of being or account of essence, with what we nowadays understand as sense. "Sense" is usually taken in the philosophical literature as expressing a or the intensional aspect of a linguistic expression's semantical value. However, tempting and didactically useful as it may be to lend "*λόγος τῆς οὐσίας*" similar contours, it must be recognised that the parallelism between the two notions is rather more minimalistic: what they indeed have in common is that both, each in its proper conceptual context, play the role of *the reason why a name applies to a thing*. It is because of the same definition of being human, or the same account of the essence of human[33], that the name "human" applies to both Plato and Socrates; and one could analogously say that Plato is human in the same sense that Socrates is human. Nevertheless, it is clear that this by no means implies that the sense of a name has anything whatsoever to do with some "essence" or "being", whatever should such notions be taken to mean; and it is unclear what the relation would be between the definition of being correspondent to a name and this name's sense, in case sense were an available notion to analyse aspects of a name in the discussed context – a point which is also unclear. From these observations, one can thus conclude that in his discussion of synonymy, and in general of the –onymies, Aristotle attaches what would nowadays be usually regarded as the semantic content or value of a name to what is probably the most crucial ontological notion within his conceptual economy, namely: *οὐσία*. It is a necessary condition for a name to apply to something that this name expresses or has associated to it a certain definition of being – or, alternatively, that this name expresses or has associated to it a certain account of an essence; and it is further required, and I assume this to be clear enough, that the οὐσία accounted for or defined *underlies* the thing in question (in the sense of "*underlie*" – a rendering of the greek expression "καθ'ὑποκεῖσθαι" [34], translated by Ackrill as "to be said of" – to be explained by Aristotle shortly after in the *Categories*, at chapter II, 1a20 – 1b10). The fact that a name applies to a thing is thus due to the concurrence of two analogous facts, namely: on the one hand, at the level of language, that the name "expresses" a certain *οὐσία*; and on the other hand, at the level of reality, that the thing "displays" such *οὐσία*.

So, just as things are synonymous in the usual sense because they express the same meaning – i.e. some semantical/linguistic feature – according to a certain criterion (say, for example, their denotation), so are things synonymous in the Aristotelian sense because they "display" the same *οὐσία* – i.e. some ontological feature – which happens to be "expressed" by a certain name. In other words: both notions of synonymy share a feature that can be seen as what is actually essential about synonymy, namely: they deem things synonymous

33 Barnes' translation seems inferior to Ackrill's in the following point: "the definition of being that corresponds to the name"(Ackrill) vs. "the account of their essence in respect of the name"(Barnes). Barnes's version, by employing the pronoun "their", explicitly attributes the "essence" (οὐσία) referred to in the phrase "account of essence"(λόγος τῆς οὐσίας) to the things candidate to being synonymous. Such explicit attribution does clearly not happen in Ackrill's version, in which the "being" (οὐσία) occurring in the phrase "definition of being" (λόγος τῆς οὐσίας) is *not* attributed explicitly to anything in particular, exactly as in the Greek text.

34 The expressions occurs in the text in present participle form, καθ'ὑποκείμενος, and has its sense contrasted by Aristotle with that of ἐν ὑποκείμενος, i.e., inherent, in a free translation. This last expression is rendered by Ackrill as "to be in".

inasmuch as they are equivalent as expressions or instances of a given principle or criterion – be the latter of an ontological or semantical/linguistic nature.

2. Synonymy and identity of proofs

Since proofs can be regarded as linguistic expressions/entities – actually, as discussed in the prologue, Aristotle's notion of *συλλογισμός* was indeed conceived in a rather linguistic register, to begin with –, it is neither surprising nor extravagant that one looks upon their *meanings* as what determines necessarily and sufficiently the behaviour of a relation of semantic equivalence holding among them; and this precise relation is a reasonable candidate to the role of what one refers to when talking about identity of proofs. This approach actually invites one to talk of *synonymy of proofs* instead of identity of proofs – a terminology that timidly showed up in some of the earliest papers to address the matter from the viewpoint of the normalisation thesis (see Kreisel 1971, p.117; Prawitz 1971, p.237), and later fell into disuse.

As already suggested in the discussion of Aristotle's notion of synonymy, contemporary use of this concept differs from that of the Stagirite in at least two important respects: it is restricted to linguistic expressions; and it seldom – if ever – incorporates a *locus* equivalent to that of the *ὄνομα* in Aristotle's version of the relation – i.e. arguably, a position for the *criterion of synonymy* .

It is not the case, of course, that contemporary synonymy is applied without any reference at all – even if implicit – to *some* criterion. Notwithstanding, the contemporary criterion of synonymy is a fixed one, the same for every case: *meaning.* Indeed, synonymy became practically a shorthand for identity viz. equivalence of meaning. This "absoluteness" of *meaning* as the criterion of synonymy could be reasonably regarded, by the way, as an explanation for the suppression of the position of criterion from the relation itself. The problem, though, is that *meaning* is itself a notion of which contemporary literature presents overwhelmingly many (and much) different accounts. The effect is that the fixation of *meaning* as the criterion of synonymy ends up being, on the one hand, enough for the suppression of the position of criterion from the relation; but, on the other, not enough for fixing just which criterion is effectively being employed from among the many possible that arise with different answers to what *meaning* means. Thus the strange appropriation and treatment of the notion of synonymy that can be witnessed in the literature of the first and beginning of the second half of the 20th century, especially in authors belonging to the so-called analytic tradition: an intense debate concerning what *the* right account (viz. criterion) of synonymy should be, intimately tied with the discussion on how the notion of analyticity should be understood.

The same conceptual atmosphere in which this sort of dispute emerged was still in force, so to speak, as the terminology of *synonymy* was dropped in the discussion of identity of proofs – which was thus presumably not a mere strike of chance. In order not to compromise oneself with a notion depicted as being rather strict (cf. e.g. Carnap's notion of *intensional isomorphism* presented in his *Meaning and Necessity* or, for an even more radical and strict picture, Goodman's "On Likeness of Meaning" and the homonymous notion he characterises therein), it is understandable that one sets apart two issues: synonymy of proofs and identity of proofs. *Synonymy* of proofs – or rather of derivations – would then be a stricter semantic question – it has to do with e.g. understanding what Carnap would have called the intensional structure of syntactic expressions of proofs; in plain terms, it is about what derivations *mean*, taken *stricto sensu*. *Identity* of proofs as conceived in the framework of the normalisation thesis, could rather be understood as a twofold issue: on the one hand, as a relation of equivalence which holds not between derivations, but rather between *what they represent* – namely proofs, however these and their representation by derivations are to be understood; on the other, in turn, again as an equivalence relation between derivations, but this time an arguably looser one than synonymy – in Goodman's terminology, it would demand a lesser degree of likeness of meaning than synonymy.

Now, in the second of its senses, identity of proofs is not necessarily distinguishable from synonymy of derivations – if so, then merely in degree, not in nature. As already said, synonymy might be regarded as stricter due to, mainly, the recent history of the term; but we could well succeed in unearthing some plausible reason to reverse the order of strictness if we were interested in fighting for these terms.

In the first of its senses, though, identity of proofs is a necessarily distinct question from any relation of semantic equivalence between derivations viz. syntactic expressions of proofs, for it concerns objects in a distinct category than that of those. This category of objects is, let us say, the most fundamental one within the context of the normalisation thesis. Just what these objects – i.e. proofs – would be, and thus what features in them should determine their identity and contribute to make them distinguishable from one another is a difficult question to answer; especially because it seems we can barely state in clear words – let alone come to an agreement about – what the subject matter is when we talk of proofs. According to how different authors approach and interpret the thesis, different objects at this level are aimed at, and thus respectively different relations of identity between them are considered to be in force: from a trivial notion of identity which only holds between an object (namely, a shared semantical content – viz. an intension or an extension – of derivations) and itself; to a relation between "intensions" which merely mirror the syntactical objects to which they are attached.

This is actually a good indication of just how astonishingly little agreement we have on the nature of such supposed objects – and there seems to be little to no sense in engaging in such a general-spirited quest for the *true* identity criteria of something about which we are, in plain terms, so clueless. A clear enough notion of proof must serve as the background of a "scientific" investigation on identity of proofs, and it comes as no surprise that any such effort is hopeless – because pointless – in its absence. This exact point is made by Troelstra in his 1975 text *Non-extensional equality* (p.318) regarding the soundness of the normalisation thesis with respect to identity of proofs:

"If we think of the objects of theory as 'direct proofs' whose canonical descriptions are normal deductions (cf. numbers and numerals in the case of arithmetic), then the inference rules $\rightarrow$I, $\rightarrow$E correspond to certain operations on direct proofs; the normalistion theorem then establishes that these operations are always defined, and permits us to regard any deduction as a description of a direct proof. The conjecture [i.e. the normalisation thesis] in the direction <= [soundness] then becomes trivially true.

But if the deductions are seen as descriptions of a more general concept of proof (...) (in the same manner as closed terms are descriptions of computations in the case of arithmetic) it may be argued that intensional equality should correspond to (literal) equality of deductions; and then the conjecture is false.

So, without further information about the intended concept of proof, the conjecture in the direction <= [soundness] is meaningless because, as just explained, ambiguous."

Such a requirement, I understand, might sound for some as an unreasonable demand for an ideal starting point – it looks like we are doing nothing but, piece by piece, starting to set the perfect scenario for a Plato-like drama consisting in Socrates torturing some poor soul into frustration by demanding an answer to the deep philosophical question: “what is a proof?”, and then saying that things shall not get started unless the whole story does not end in aporia. As far as we are concerned, there is actually no reason to assume that such a quest would have any point at all; it might well be that there is no such thing as THE notion of proof, the nature of which is lying in the deep shadows of oblivion, waiting to be discovered and revealed by some chivalrous *εἰδότα φῶτα*. But it should be made clear from the very beginning that the setting of the starting point of a reasonable attempt to deal with the question of the identity of proofs need neither involve such an epic *“τί ἐστι”* enterprise, nor be philosophically conformist or arbitrary. Indeed, all the information we need is at everyone’s disposal, and it is namely this: we use proofs and talk about them in various contexts; and both our use and talk of proofs should provide us with enough indication to, in each proper context, properly carry out the evaluation of any attempted answer to questions regarding the identity of proofs. Generality does not have to be an aim.

Ironically enough, there are also those who, in an attitude that might sound quite contrary to the remarks just made, claim that identity of proofs is an issue of the utmost importance in answering the question “what is a proof?” (see e.g. Kosta Došen 2003). This is of course not wrong; without the determination of a proper criterion of identity, nothing like the notion of *a proof* (as opposed to other proofs of the same kind, so to say) could be achieved. But there is a difference between this viewpoint and that according to which lack of clarity on what proofs are prevents the investigation of their identity: the latter aims at *describing* proofs, while the second is a *definitional* approach to proofs. While these two endeavours can of course be synchronised and even carried out by means of a single effort, they obviously do not have to be; and it is quite hard to evaluate to which extent or in which sense the normalisation thesis succeeds in the task of carrying either of them out.

Let us now make some considerations on how the normalisation conjecture compares to other compelling ideas concerning synonymy of derivations and identity of proofs.

a. The normalisation thesis and a notion of synonymy of proofs inspired by Carnap’s intensional isomorphism

Firstly, we will define a notion of synonymy of proofs which is in tune with the *Zeitgeist* of the investigations of the first half of the 20^{th} century to serve as a measure of evaluation of the normalisation thesis in two senses: in the first place, as itself another notion of synonymy of proofs intended at capturing a stricter intensional content of derivations – if this makes any sense –, and in the second, as a formalisation of a notion of identity of proofs.

A preliminary notion is thus needed:

<u>Definition of rule isomorphism</u>:

Given two inference rules belonging respectively to the connectives $*^n$ and $*'^m$, let A_i resp. A'_j, $B_{i'}$ resp. $B'_{j'}$, $\Gamma_{i''}$ resp. $\Gamma'_{j''}$, C resp. C' be, respectively: a (minor) premiss of $*^n$ rule resp. $*'^m$ rule; an argument expression of the occurrence of $*^n(B_1,\ldots,B_n)$ resp. $*'^m(B'_1,\ldots,B'_m)$ in the $*^n$ rule resp. $*'^m$ rule; a set of assumptions on which a (minor) premiss of $*^n$ rule resp. $*'^m$ rule depends; and the conclusion of the of $*^n$ rule resp. $*'^m$ rule iff $*^n$ rule resp. $*'^m$ rule is an elimination rule. $*^n$ rule and $*'^m$ rule are said to be *isomorphic* if and only if there’s a bijective correspondence f between their aforementioned syntactic elements such that:

a) $f(A_i) = A'_j$
b) $f(B_{i'}) = B'_{j'}$
c) $f(\Gamma_{i''}) = \Gamma'_{j''}$
d) $f(C) = C'$
e) $f(A_i) = A'_j$ iff $f(\Gamma_i) = \Gamma'_j$
f) If $f(\Gamma_{i''}) = \Gamma'_{j''}$, then $\Gamma_{i''}$ and $\Gamma'_{j''}$ have the same number of elements.
g) (Anaphora preservation) For each A_i $[B_i]$ $[D_q \in \Gamma_i]$ $[C]$, if and only if A_i $[B_i]$ $[D_q \in \Gamma_i]$ $[C]$ is identical with, respectively, either $A_{i'}$ or B_i' or $D_{q'} \in \Gamma_{i'}$ or C, then $f(A_i)[f(B_i)]$ $[D'_k \in f(\Gamma_i)]$ $[f(C)]$ is identical with, respectively, either $f(A_{i'})$or $f(B_i')$ or $D'_{k'} \in f(\Gamma_{i'})$or $f(C)$.

Remarks:

i. The schematic variables A, B, C and D (with or without primes or indexes) in the rule schemata may only be instantiated by *atomic* schematic variables; in other words, the only inference rules considered are introduction rules in which there are no occurrences of operators but that mentioned in the conclusion, and elimination rules in which there are no occurrences of operators but that mentioned in the major premiss. In the *applications* of the inference rules, on the other hand, other operators may occur accordingly;
ii. An operator has a *set* of inference rules, i.e. it cannot have two or more identical inference rules associated to it.

We now proceed to the definition of the notion itself:

Definition of synonymy ($\equiv_{int}$):

1. For atomic A and B: $A \equiv_{int} B$ iff $A \vdash B$ and $B \vdash A$;

2. For connectives: $*^n \equiv_{int} *'^m$ iff for each rule of $*^n$, there is a rule of $*^m$ (and vice-versa) such that the two rules are isomorphic;

3. $A := *^n(A_1,\dots,A_n)$ and $B := *'^m(B_1,\dots,B_m)$: $A \equiv_{int} B$ iff:
   a) $*^n \equiv_{int} *'^m$;
   b) there is a bijective function $f: A_i \mapsto B_j$, $0 \leq i \leq n$, $0 \leq j \leq m$, such that $f(A_i) \equiv_{int} A_i$;

4. $\Gamma := \{A_1,\dots,A_n\}$ and $\Delta := \{B_1,\dots,B_m\}$: $\Gamma \equiv_{int} \Delta$ iff there is a bijective function $f: A_i \mapsto B_j$, $0 \leq i \leq n$, $0 \leq j \leq m$, such that $f(A_i) \equiv_{int} A_i$;

5.For derivations:
5.1. Let Π and Π' be , respectively, the following derivations ending in introductions:

$$\frac{\begin{matrix}[\Gamma_1], & \dots, & [\Gamma_n] \\ \Sigma_1 & & \Sigma_n \\ A_1, & \dots, & A_n\end{matrix}}{*(B_1,\dots,B_n)} \quad \text{and} \quad \frac{\begin{matrix}[\Gamma'_1], & \dots, & [\Gamma'_m] \\ \Sigma'_1 & & \Sigma'_m \\ A'_1, & \dots, & A'_m\end{matrix}}{*(B'_1,\dots,B'_m)}$$

Then $\Pi \equiv_{int} \Pi'$ iff:

a) for every derivation Σ_i of premiss A_i depending on Γ_i, there is a derivation Σ'_j of premiss A'_j depending on Γ'_j such that:

(i) $\Gamma_i \equiv_{int} \Gamma'_j$;

(ii) $A_i \equiv_{int} A'_j$;

(iii) $\Sigma_i \equiv_{int} \Sigma'_j$; and

b) for every derivation Σ'_j of premiss A'j depending on Γ'j, there is a derivation Σ_i of premiss A_i depending on Γ_i such that:

(i) $\Gamma_i \equiv_{int} \Gamma'_j$;

(ii) $A_i \equiv_{int} A'_j$;

(iii) $\Sigma_i \equiv_{int} \Sigma'_j$; and

c) $*(B_1,\ldots,B_{n'}) \equiv_{int} *'(B'_1,\ldots,B'_{m'})$;

5.2. Let Π and Π' be, respectively, the following derivations ending in eliminations:

$$\frac{\begin{matrix} [\Gamma_1], & \ldots, & [\Gamma_k] & & [\Gamma_q], & \ldots, & [\Gamma_{n'}] \\ \Sigma_1 & & \Sigma_k & & \Sigma_q & & \Sigma_{n'} \\ A_1, & \ldots, & A_k, & *^n(B_1,\ldots,B_n), & A_q, & \ldots, & A_{n'} \end{matrix}}{C}$$

and

$$\frac{\begin{matrix} [\Gamma'_1], & \ldots, & [\Gamma'_{k'}] & & [\Gamma'_{q'}], & \ldots, & [\Gamma'_{m'}] \\ \Sigma'_1 & & \Sigma'_{k'} & & \Sigma'_{q'} & & \Sigma'_{m'} \\ A'_1, & \ldots, & A'_{k'}, & *'^m(B'_1,\ldots,B'_m), & A'_{q'}, & \ldots, & A'_{m'} \end{matrix}}{C'}$$

Then $\Pi \equiv_{int} \Pi'$ iff:

a) $*(B_1,\ldots,B_{n'}) \equiv_{int} *'(B'_1,\ldots,B'_{m'})$;

b) for every derivation Σ_i of minor premiss A_i depending on Γ_i, there is a derivation Σ'_j of minor premiss A'j depending on Γ'j such that:

(i) $\Gamma_i \equiv_{int} \Gamma'_j$;

(ii) $A_i \equiv_{int} A'_j$;

(iii) $\Sigma_i \equiv_{int} \Sigma'_{j;}$

c) for every derivation Σ'_j of minor premiss A'j depending on Γ'j, there is a derivation Σ_i of minor premiss A_i depending on Γ_i such that:

(i) $\Gamma_i \equiv_{int} \Gamma'_j$;

(ii) $A_i \equiv_{int} A'_j$;

(iii) $\Sigma_i \equiv_{int} \Sigma'_{j;}$

d) $C \equiv_{int} C'$

a.1. Remarks on this notion of synonymy

Those familiar with Carnap's 1947 work *Meaning and Necessity* will probably have recognised that the notion of synonymy, as well as the notion of *meaning* that it inevitably suggests, are not but an adaptation of an original idea of his. They correspond – and not really roughly – to his notions of *intensional isomorphism* and *intensional structure* (cf. §14, chapter I of the mentioned book). Notwithstanding, there are some noteworthy features of the present account of this idea that differ from Carnap's own exposition of it. The three first are obvious enough:

(a) Carnap's own treatment of the idea does not contemplate derivations, while the one offered here has been formulated purposely to give an account of them;

(b) he defines the notion in a much broader way – namely, for any language whatsoever which has a certain grammatical structure. Here, in turn, a more specific – though comprehensive enough – formalism is addressed: natural deduction systems for propositional logic, with the due mentioned restrictions; and

(c) while Carnap relies on model-theoretic semantic notions to formulate his idea, it is rendered here in a purely proof-theoretic semantic framework – i.e. instead of basing synonymy in notions such as L-equivalence, the notion of interderivability and a simple correspondence between the syntactic structure of the inference rules of propositional operators are used as touchstone.

There are also some interesting differences which concern, I believe, the motivation behind the two formulations. It seems, for instance, that Carnap put the possibility of comparing the meaning of expressions formulated in different languages in foreground when providing his definitions and explanations on the matter. While composing the present account, on the other hand, no attention whatsoever was paid to this. It is simply focused on defining a rudimentary notion of logical synonymy for some given language/deductive system which displays the required features. Because of this peculiarity of Carnap's aim, I daresay that his own presentation of the idea ends up lending it outlines of the "*explicatum"* – to use a word of his – of a hybrid between synonymy and *translation*; notions which are, of course, quite different in relevant ways. It seems one could provide sensible reasons to support that, say, (A&B) is a good translation of (A ^ B) while (B&A) is not; after all, as Carnap himself puts it, meaning components other than those which he calls "cognitive" or "designative" are of capital relevance for the adequate translation of sentences. On the other hand, it is likely that a *logical* account of synonymy which entails – as Carnap's own does – that, for instance, it is a necessary condition for having non-homonymous synonymous expressions belonging to a same given language that there is some atomic component of one which is not a component of the other – which implies that e.g. the *meaning* of (A&B) differs from that of (B&A) – would be contrary to some familiar and compelling pre-conceived ideas on the matter.

As concisely as I am capable of putting it, it seems to me that Carnap is not clear enough with respect to his choice of criteria to govern his use of notions such as syntactical structure and intensional structure. In fact I believe he was specially unfortunate in formulating rule *b* of his definition of intensional isomorphism – one of whose clauses states that it is a necessary condition for two expressions A and B to be intensionally isomorphic that their correspondent component expressions be not only intensionally isomorphic, but also occupy the same position as arguments in the main expression. On this specific rule of his definition, he comments:

"In accord with our previous discussion of the explicandum, rule b in this definition takes into consideration the order in which argument expressions occur but disregards the place of the main subdesignator. For the intensional structure, in contrast to the merely syntactical structure, only the order of application is essential, not the order and manner of spelling."

A page later, in the beginning of §15, he remarks a striking similarity between the notion of intensional isomorphism just outlined and that of *equivalence of analytic meaning* put forward by C. I. Lewis. He quotes Lewis's definition of the notion, and then goes on to point at some differences between it and his own account of synonymy – which leads him to make some critical comments on Lewis that, although quite to the point, are not exactly relevant for our present considerations. Curiously enough, though, Carnap does not mention a relevant difference between the mentioned clause of his rule b and the "correspondent" clause in Lewis's definition. The requirement made by Lewis's clause is that "the order of corresponding constituents is the same in both [expressions], <u>or can be made the same</u> without alteration of the intension of either whole expression" (emphasis by myself) – a clearly less strict one than Carnap's own. It allows, for instance, for non-homonymous expressions of a certain language with exactly the same atomic components to be synonymous – in particular, it entails that e.g. (A&B) and (B&A) are synonymous[35]. This might be regarded as an advantage of Lewis's formulation over Carnap's. Of course, the "order of application" is, generally speaking, essential to the intensional structure; one could not claim with no further that e.g. (A→B) and (B→A) are even extensionally, let alone intensionally the same. It just does not follow from this that *any* difference whatsoever in the order of application will yield different intensional structures. Why should we regard the order of application as essential to the *meaning* of e.g. a propositional operation such as conjunction? This is a point where the motivational difference between Carnap's exposition and the present, as mentioned above, shows some of its consequences.

If this series of remarks on the matter were to stop now, it would sound as if it were being suggested here that Carnap has depicted with his stricter definition of intensional isomorphism – his "*explicatum*" for synonymy, let us keep in mind – an equivalence relation looser than identity of what he calls syntactical structure and yet stricter than identity of intensional structure, the one which he actually intended to characterize. In crescent order of strictness, we would have: (i) identity of intensional structure; (ii) the equivalence relation actually defined by Carnap as intensional isomorphism; (iii) identity of syntactical structure; (iv) syntactical identity or homonymy, i.e. to be a copy of; (v) sameness as an occurrence of a given sign. Relation iii – or, better said, the notion of syntactical structure, in terms of which it is formulated – is left unspecified by Carnap to the point of, I believe, jeopardising its

35 Strictly speaking it does not entail this, because Lewis's defines the relation "equivalence of analytic meaning" only for valid and inconsistent statements. This restriction can nevertheless be safely – and quite profitably, as Carnap himself suggests – ignored.

meaningfulness. Relation ii in turn, though well explained enough, has also no clear role in the discussion of synonymy; perhaps some adequate variation of it could yield good fruits if employed for explaining aspects of the concept of *translation* from a logical viewpoint.

There is nevertheless one more aspect worth mentioning; actually, something that Carnap wrongly dismisses as unproblematic, but is a quite relevant topic. When comparing, for example, expressions like ((A&B)&C)and its counterpart &&ABC in prefix notation, it seems that he would have supported that they should be regarded as intensionally isomorphic. This position is curiously loose, though – it just seems that infix notation can generate ambiguous "expressions" which are simply impossible to construct in prefix notation, where parentheses are unnecessary. Of course, one could object to this by reminding me of the reason why I put the word "expressions" between quotation marks in the previous sentence – i.e., there are no ambiguous expressions in infix notation because what *would* be ambiguous expressions is simply ruled out by an additional grammatical stipulation: the compulsory parenthisation of compound (sub)expressions – including of some to which such a move represents no change whatsoever in meaning. But in prefix notation it is just so that one could not possibly abolish some disambiguation rule or convention in order to allow ambiguous expressions to occur – ambiguity is ruled out from scratch, without redundancies, irrelevancies (e.g. outermost parentheses) or priority conventions, by the very grammatically determined syntactical structure of compound expressions. In other words, there is a grammatical incongruence between languages with infix and prefix notation which suggests a difference in intensional structure between expressions one would regard – quite correctly – as *translations* of one another in two such languages. No such incompatibility would occur, though, should we compare expressions in languages with prefix and postfix notation – a case in which only a difference in "manner of spelling" would in fact obtain.

Three morals to this last part of the story: (a) identifying, even if only extensionally, the notions of "cognitive" and "designative" meaning – or, put another way, understanding meaning insofar as it concerns logic in terms of designation – apparently lead Carnap to the impossibility of distinguishing some shades of grey between intension and syntax that are important for the adequate understanding of a notion such as intensional structure, for instance, subtle intensional differences such as the one just described between infix and prefix notations; (b) even from a logical standpoint, translation is a conceptually troublesome business, which requires the preservation of sometimes less, sometimes more than synonymy – which is all by itself a notion surrounded by enough mist; (c) one should use extreme doses of both caution and honesty when employing technical vocabulary and devices for clarifying conceptual purposes. All understand perfectly that parentheses are disambiguation devices; therefore, there is no point in saying that things such as A&B&C are devoid of meaning only because grammar stipulates that they are not expressions or well-formed formulas – ambiguous and meaningless are, after all, at least quite different if not indeed contrary notions. Actually, one could regard precisely this sort of "expression" – "ill-formed" formulas, if you will; something whose distinction from things such as, say, &A→˅B is quite clear and precise – as a good reason why people insist on adding the frequently referred to as meaningless "well-formed" before "formulas".

a.2. Comparison between the presented notion of synonymy and identity (or synonymy) of proofs according to the normalisation thesis

Some expectations as to how a notion of synonymy in the spirit of the one defined here and a notion of equivalence of derivations aimed at formalising identity of proofs should behave with respect to each other seem to make sense. Following the considerations of Carnap and C.I. Lewis, the notion of synonymy defined above is very strict and somewhat dull – its goal is to capture, so to speak, the strictest level of intensional content expressed by an expression which is wider than its mere syntactical features. It is thus expected that any other significant semantical equivalence relation between derivations – be it thought as intensional or extensional – is to be by necessity complete with respect to this sort of relation of synonymy. Such an expectation is however obviously not met by means of the adoption of the notion of synonymy just defined and of a notion of equivalence of derivations in accordance with the normalisation thesis.

THEOREM: there are non-βη-equivalent synonymous derivations.
PROOF: There are synonymous derivations which do not share the same end-formula and hence cannot be βη-equivalent, e.g.:

$$\frac{p}{p \vee q} \qquad \frac{p}{q \vee p}$$

Q.E.D.

If we accept the notion of synonymy just presented and, simultaneously, that the equivalence relation between derivations yielded by the normalisation thesis is a good formal interpretation of identity of proofs, then we are forced to the conclusion that the latter is, quite counter-intuitively, not complete with respect to synonymy of derivations. Whether we take identity of proofs as a relation of extensional or of intensional nature, it will be expected that synonymous derivations – in the strict Carnap-Lewis sense intended – must be/express/represent the same proof. One could, of course, say that the notion of synonymy presented is too coarse and should not allow for non-βη-equivalent derivations to be synonymous – but then again, it seems that synonymy will have to be thought as coinciding with plain syntactical identity. On the other hand, in turn, if one is inclined to accept that all derivations deemed synonymous by the notion just presented indeed share the sort of intensional content aimed at by Carnap and others, it seems that we would have to incorporate some more derivations to the list of those which are identified by the normalisation thesis as being/expressing/representing the same proof – namely all non-βη-equivalent synonymous derivations. The two derivations exemplified above, for instance, again point at a taxonomical issue involved in the normalisation thesis: the fact that there can be no two equivalent derivations with different end-formulas makes us wonder, whether we take the normalisation criterion as restricted to the result or not, about the extent to which the normalisation thesis can be understood as compatible with a general, non-trivial picture of synonymy of formulas, for instance.

The situation is nevertheless quite obviously worse than what the derivations just exemplified show. Thus:

THEOREM 2: there are non-βη-equivalent synonymous derivations sharing the same set of assumptions and end formula.

PROOF:

$$\frac{p \wedge p}{p}{\scriptstyle \wedge E1} \qquad\qquad \frac{p \wedge p}{p}{\scriptstyle \wedge E2}$$

Q.E.D.

Synonymy as defined here thus additionally identifies some derivations *from a certain Γ to a certain A* which are not identical according to the normalisation thesis. Because of the already mentioned Post-completeness results, however, this means that adding all synonymous derivations to the list of those identified by the normalisation thesis will trivialise the identity of derivations from Γ to A for at least some Γ and some A which yielded more than one proof. This clearly poses an even stronger argument for the incompatibility between the present notion of synonymy and the notion of identity of the normalisation thesis: for it is not just that the latter is merely incomplete with respect to a dully strict notion of synonymy such as the former; it also simply cannot be completed as to encompass the dully strict idea of the former without significant losses. This could also serve as a good example of how the results of Post-completeness concerning the notion of identity yielded by the normalisation thesis can be regarded as showing a weakness rather than a merit of it.

Notwithstanding, there are limits to the damage caused by the addition of synonymous derivations to those identified by the normalisation thesis. For the next theorem, we need preliminary notions:

a)Counterpart function: let Π and Π' be synonymous derivations. A function Cp: Π --> Π' is called a *counterpart function* iff:

(i) Cp(Π) = (Π');

(ii) let Σ be a subderivation of Π such that the distance of the end formula of Σ from the end formula of Π is n. Then Cp(Σ) = Σ', where Σ' is a subderivation of Cp(Π), Σ' $\equiv_{int}$ Σ and the distance of the end formula of Σ from the end formula of Cp(Π) is n;

(iii) let A be the end formula of a subderivation Σ of Π. Then Cp(A) = A', where A' is the end formula of Cp(Σ);

(iv) Let B be a subformula of the end formula A of a subderivation Σ of Π. Then Cp(B) = B' , where B' is a subformula of Cp(A) such that B $\equiv_{int}$ B'.

b) Perfect synonymy: Let Π and Π' be derivations. Π and Π' are *perfectly synonymous* iff there is a counterpart function Cp: Π --> Π' such that Cp(Π) [Cp(Σ)] [Cp(A)] [Cp(B)] = Π [Σ] [A] [B].

In plain words, two synonymous derivations are perfectly synonymous iff they share exactly the same formulas in the same "places"!

THEOREM 3: let Π and Π' be synonymous derivations and share the same set of hypotheses Γ and end formula D; then the normal forms of Π and Π' are perfectly synonymous.

PROOF:

- If Π and Π' are perfectly synonymous, the theorem follows trivially;

- Suppose now that Π and Π' are not perfectly synonymous. Then there are occurences of formulas A and B such that $A \neq B$, $A \equiv_{int} B$, A[B] occurs in Π[Π'], and there is no Cp: Π --> Π' such that Cp(A)=A;

- Let now all $\Pi_i[\Pi'_i]$ (i<n) be reducts of Π[Π'] and $\Pi_n[\Pi'_n]$ the normal form of Π[Π']. Then either (a)A[B] occurs as a maximal formula in some $\Pi_i[\Pi'_i]$, or (b) A[B] occurs in $\Pi_n[\Pi'_n]$;

- In case (a), then A[B] does not occur in $\Pi_n[\Pi'_n]$, which implies that Π_n and Π'_n are perfectly synonymous;

- In case (b), then, by the subformula principle, either A[B] is a subformula of D, or A[B] is a subformula of some C ∊ Γ. But it cannot be the case that A is a subformula of either D or of some C ∊ Γ – were it the case, then, since Π and Π' share D and Γ, there would be some Cp: Π --> Π' such that Cp(A)=A, which contradicts the hypothesis under which we are that there is no Cp: Π --> Π' such that Cp(A)=A. Since (b) takes us to contradiction, then A[B] does not occur in $\Pi_n[\Pi'_n]$; which again implies that Π_n and Π'_n are perfectly synonymous;

- Therefore, Π_n and Π'_n are perfectly synonymous.

Q.E.D.

Together with theorem 2, theorem 3 tells us that all non-equivalent derivations from a certain Γ to a certain A identified by synonymy reduce to perfectly synonymous normal forms. In other words, given the maximality results, if one accepts the normalisation thesis, theorem 3 tells us *which* derivations have their identity trivialised by synonymy. Thus, with the acceptance of the normalisation thesis, synonymy trivializes the identity of all and only those derivations from a certain Γ to a certain A such that there are at least two non-equivalent perfectly synonymous normal derivations Π and Π' from Γ to A.

b. Comparison between the normalisation thesis and the thesis of "principality"

Another interesting way of approaching identity of proofs is by means of an idea suggested by the notion of principal type of a λ-term. In a natural deduction setting, the notion of principal type of a term corresponds to the end-formula of a maximally generalised derivation. This idea was already explored in III.2.a.1 and it is considered in some works concerning identity of proofs – though not itself as a supposedly sound and complete criterion for identifying derivations that correspond to identical proofs (see e.g. Widebäck 2001). As a full thesis on identity of proofs, it could be basically expressed like this: two derivations are equivalent as proofs iff they have the same principal type, i.e. iff the conclusions or their maximal generalisations are the same. Let us call this the principality thesis on identity of proofs.

Inversely to what (for some) happens with the normalisation thesis, the appealing part of the principality thesis is supposedly its claim to completeness with respect to identity of proofs, its soundness being more obviously objectionable. The reason for this situation is basically the same sort of interpretation that led Kosta Došen to see in the relation between principality and β- and η- conversions a reason to raise an objection against the soundness of the normalisation thesis: in the first place, the idea that the principal type of a derivation, not its end-formula, is the actual result of the proof at stake in it; and secondly, that all identical proofs must share the same result. Since the principality thesis does identify all and only derivations with the same principal type, it thus generates a class of equivalence of derivations all of which certainly satisfy a supposedly *necessary* condition for representing identical proofs – i.e. sameness of result –; thence the trustworthiness of the claim to completeness. Soundness, on the other hand, must be somehow further argued for.

Another way of looking at the meaning of generalising a derivation and the notion of principal type within the framework of identity of proofs and the principality thesis was suggested to me by Prof. Schroeder-Heister: if one erases the formulas of a derivation, but keeps the applications of inference rules and discharge of hypotheses untouched, one obtains a sort of formal skeleton of the proof – which corresponds to an untyped lambda-term. Thanks to the principal type theorem, we know that such structures can be assigned a principal type viz. have a maximally generalised version. Thus, the conclusion of a maximally generalised derivation – which corresponds to the principal type of a lambda term – is the expression of the *potential* of such argumentative structure, viz. the most general result it is capable of proving. The claim involved in the principality thesis would then be quite different: sameness of potential understood in this sense would be the necessary and sufficient condition for two derivations to be deemed equivalent as proofs.

Understood in this second way, the principality thesis seems to pose a similar question regarding its soundness in case of completeness: it relies on the notion of potential of a skeleton defined by means of the notion of principal type and stipulates thereby a notion of equivalence such that derivations whose skeletons share the same potential i.e. the same principal type should be identified; but it seems that, if anything, sameness of potential should be a necessary rather than a sufficient condition for two derivations to be the same as proofs.

The difference between these two approaches lets itself be explained by, again, a taxonomic issue. The first approach, on the one hand, identifies proofs for having the same *result*; the second, int turn, for having the same *potential*. So, while the first could be described as regarding proofs as *performed* deductions considered *in abstracto* – thus putting proofs in the category "performed" –, the second seems to understand proofs as deductions *that can be performed* – thus framing proofs as belonging to the category "performable". In this way, by understanding proofs as respectively different things, what the first approach regards as the *result* of a proof is exactly what the second regards as the *potential* of a proof.

As observed in the comparisons between derivations made in section III.2.a.1, neither β- nor η- conversions preserve principality – which means that normalisation thesis identifies derivations with different principal types. Therefore, no matter in which of the two ways described above principality thesis is understood, it seems to be in direct conflict with the normalisation thesis in the following way: its completeness, for which we arguably have convincing informal grounds, implies that the soundness of the normalisation thesis, for which we also purportedly have equally convincing informal grounds, does not hold. And the converse is also the case: if normalisation thesis is sound, then principality thesis is not complete. This means that the two basic intuitions about proofs and their identity which we supposedly have enough and equally strong reasons to accept from the outset are incompatible. Which one would then be right?

Such a dilemma is in fact the fruit of a misunderstanding. The reason for it to come about is fundamentally the idea that both the normalisation thesis and the principality thesis are incompatible competing ways to answer to the same question, namely, the one concerning the criteria of identity of proofs, when in fact they are not – or at least not in a way that poses any significant problem. The actual incompatibility is indeed a formal one: because the normalisation thesis is not sound with respect to the principality thesis, they cannot both hold as, respectively, sound and complete criteria of equivalence of derivations. But they can both be taken to hold as criteria for identity of proofs, for the proofs "individuated" by each of such criteria of identity could simply be taken to, say, belong to different types. So, just as <2,3> and <4,6> are the same rational number but not the same ordered pair of natural numbers, so are the pairs of derivations in the examples in section III.2.a.1 the same proof – when proof means what is individuated by the identity criterion yielded by the normalisation thesis and they are taken to be proofs – and also not the same proof – when proof means what is individuated by the identity criterion yielded by the principality thesis and they are taken to be proofs. One could, it is true, wonder which of the two types, if any, *truly* corresponds to *THE concept* of proof; but, as already suggested before, as far as we are concerned, such a thing is most likely a senseless idea.

But in any case, there is something that attention to our taxonomy of criteria may tell us concerning the adequacy of each of these proposals. While normalisation thesis seems in no sense to fit into any of the "trivial" categories, the principality thesis seems clearly to be intended as a criterion which is trivial and restricted to the result. For it yields a criterion that reduces the identity of proofs to that of something different – in this case, either the result or the potential of the structure of a derivation. In this sense, despite yielding a syntactically non-trivial equivalence relation between derivations (even when one considers derivations from a certain Γ to a certain A) just as much as the normalisation thesis does, it is a way of putting proofs apart that essentially fits into the "purely extensional" background conception that two proofs are equal iff they have the same result. From the viewpoint that motivates the normalisation thesis, such a proposal would thus be, roughly put, a mere syntactical sophistication of the idea *against which* the investigation of identity of proofs started to be undertaken in the first place, namely: that there can be no two different proofs of the same result. So, as indicated by our taxonomic evaluation, the principality thesis yields a *trivial* criterion of identity of proofs *restricted to the result*. And this is definitely an important information when considering how this criterion shapes our understanding of what a proof is/means in contexts in which it is adopted, and also which tasks/questions such a criterion is able to perform/answer.

V. Towards an evaluation of the normalisation thesis: the case of Church-Turing thesis as touchstone

In order to properly assess whether or not a given enterprise was successful, it is usually required that one knows clearly the purpose for which it has been undertaken, or according to which it is to be evaluated. In the eventual absence of such a purpose, any question concerning success or failure becomes unanswerable, because ultimately senseless. And in case the purpose is vague and unclear, or in case there are many distinct, sometimes even conflicting purposes attached to the enterprise, then the judgement of its success is doomed to be, even ideally considered, equally vague, unclear or conflicting. These quite unremarkable observations should suffice to convey to the reader the core reason why it is ultimately impossible that the comments to be made in this section represent any sort of proper evaluation of the success of the normalisation thesis itself: for the purpose or purposes for which it has been put forward or according to which it should be evaluated have not been stated explicitly or clearly enough, and are not really well understood or agreed upon as yet.

This is hardly the first occasion where this sort of observation is made regarding the normalisation thesis. Troelstra, for instance, in a passage already transcribed here[36], addresses the very same issue: there, he calls the thesis ultimately “meaningless”, because “ambiguous” with respect to the notion of proof intended – i.e. since it is not possible to determine which notion of proof the thesis aim to describe, its merit could not be evaluate. The formulation of his diagnosis does make it sound senseless, though; for ambiguity comes about not where meaning lacks, but rather contrarily, where there is a kind of superabundance thereof. It is also not quite clear if it is due to ambiguity, vagueness or some other phenomenon concerning the terms and concepts involved that we lack reasons to properly judge how well the normalisation thesis fares overall. Unlike Troelstra, however, we shall here not simply rest contented with stating the impossibility of making a precise evaluation of the merits of the normalisation thesis due to our difficulty in determining its actual purposes. There is, after all, quite a number of relevant and clear enough philosophical projects which one could think such a thesis might be directed at, or maybe not unreasonably considered as a candidate to adequately perform – and thus we can, supposition after supposition, come to a better understanding of, at the very least, the extent to which some of the philosophical

36 Section IV.2.

projects normalisation thesis is at least in principle capable of performing are worth pursuing. In what follows, a comparison will be made between the normalisation thesis and the Church-Turing thesis concerning some key aspects of them. Due to their several and significant conceptual similarities, as well as convenient differences, and to the possibility of a more solid historical assessment of the latter, this will be useful to give us a measure for the reasonability of our eventual judgements and expectations concerning the former.

1. The Church-Turing thesis and the normalisation thesis compared

The famous epithet *Church-Turing thesis* is employed in the literature to refer to a variety of related theses that typically characterise some notion of *computation* or *effectiveness* by means of the formal notions of Turing machines or λ-definability. Here, we will use it to primarily refer to the following formulation: All and only *effective computations* can be carried out by a Turing machine viz. are λ-definable. On some occasions, the normalisation thesis has been deemed analogous to the Church-Turing thesis with respect to how it addresses its object. Kosta Došen 2003, p.4, for instance, says:
"The Normalization Conjecture is an assertion of the same kind as Church's Thesis: we should not expect a formal proof of it. (...) The Normalization Conjecture attempts to give a formal reconstruction of an intuitive notion. (Like Church's Thesis, the Normalization Conjecture might be taken as a kind of definition. It is, however, better to distinguish this particular kind of definition by a special name. The Normalization Conjecture (...) might be taken as a case of analysis (...))"

a. Other theses more closely analogous to Church-Turing thesis

To be more accurate, the analogy of Church-Turing thesis seems to be stricter with other theses put forward within the literature of general proof theory, closely related to the normalisation thesis yet distinct from it. One, for instance, is claimed by Prawitz 1971, and could be stated as follows: every proof in first order logic can be carried out by means of some Gentzen-style natural deduction derivation, and by means of every such derivation can some such proof be carried out. Indeed, one can see that Prawitz underlines still a further point of analogy between the specific formulation of Church-Turing thesis in terms of Turing machines and the thesis he claims in his text in terms of Gentzen-style natural deduction derivations, namely: both these formalisms would yield “completely analysed” versions of what they render formally. Such analogy is explicitly stressed by Prawitz in his text (1971, 2.1.3, p. 246):

“(...)Gentzen’s systems of natural deduction are not arbitrary formalizations of first order logic but constitutes a significant analysis of the proofs in this logic. The situation may be compared to the attempts to characterize the notion of computation where e.g. the formalism of μ-recursive functions or even the general recursive functions may be regarded as an extensional characterization of this notion while Turing’s analysis is such that one may reasonably assert the thesis that every computation when sufficiently analysed can be broken down in the operations described by Turing.”

Dummett’s so called “fundamental assumption”is also more strictly analogous to Church-Turing thesis than the normalisation thesis. In Dummett’s own words, the fundamental assumption states that ‘‘if we have a valid argument for a complex sentence, we can construct a valid argument for it which finishes with an application of one of the introduction rules governing its principal operator’’ (Dummett 1991, p.254) – i.e. a valid argument in canonical form. Turing 1954 puts forward a thesis closely related to Church-Turing thesis and with respect to the nature of puzzles that illustrates this analogy most clearly:

“(…) the *normal form for puzzles is the substitution type of puzzle*. More definitely we can say: *Given any puzzle we can find a corresponding substitution puzzle which is equivalent to it in the sense that given a solution of the one we can easily use it to find a solution of the other*.”

Where Turing speaks of puzzles in this passage, one could well speak of effective computations, valid arguments or proofs; where he speaks of the substitution type of puzzle, one could talk in terms of, respectively, Turing machines or valid canonical arguments; and where he speaks of a correspondence viz. equivalence between solutions, one could speak of sameness of, respectively, inputs and outputs or assumptions and conclusion.

Realise, further, that the status of such theses depends crucially on how one approaches the notion characterised by the thesis in question. If one, for instance, takes the valid arguments referred to by the fundamental assumption to be natural deduction derivations in intuitionistic first order logic, the “assumption” becomes in fact a *theorem* – indeed, a corollary to the normal form theorem. A more general or a less sharp viz. formally stated view of what one understands by *a valid argument for a complex sentence*, however, may lend the assumption more *definitional* contours. In fact, this is exactly what Turing observes about the above stated thesis on puzzles, right after formulating it:

“This statement is still somewhat lacking in definiteness, and will remain so. I do not propose, for instance, to enter here into the question as to what I mean by the word ‘easily’. (...) In so far as we know a priori what is a puzzle and what is not, the statement is a theorem. In so far as we do not know what puzzles are, the statement is a definition which tells us something about what they are.”

We will come back to this passage (probably still a few times) later, for it also applies in important senses to the normalisation thesis.

2. First relevant point of analogy: characterisations of a notion in formally tractable terms which claim to be sound and complete

Notwithstanding the observations just made, the analogy observed by Došen between the normalisation thesis and the Church-Turing thesis is still significant and illuminating. In fact, all of the above mentioned theses share a very important trait with the normalisation thesis, namely: all theses offer a characterisation of a notion – say, an “*explicandum”* – in terms of a formally tractable concept, and claim that this formal characterisation is sound and complete with respect to the characterised notion. Thus, just as the Church-Turing thesis suggests the possibility of soundly and completely characterising the informal notion of effective computability at stake in e.g. Turing’s 1936 paper[37] in terms of the formal notions of Turing-machines and λ-definability; just as Prawitz 1971 characterises first order logic proofs as soundly and completely[38] analysable in terms of natural deduction derivations; just as

37 Which is a quite specific one; more on this topic later.

38 This “completely” holds both in the sense that all proofs can be analysed in this way and in the sense that the analysis is complete i.e. breaks them down into their most elementary component steps; yet the first of these senses is the obviously the relevant one to the described analogy.

Dummett assumes that the existence of canonical valid arguments of appropriate form is necessary and sufficient to explain the validity of arguments in general; and just as Turing proposes that the specifically defined notion of substitution puzzle soundly and completely characterises the general notion of puzzle; so does the normalisation thesis intend to provide a sound and complete characterisation of the informal viz. semantical notion of identity between proofs viz. semantical equivalence between derivations in terms of the formal viz. syntactical notion of (βη-)equivalence based on the reductions involved in the normalisation of natural deduction derivations. By means of the normalisation thesis, though, another thesis that is again closer in form to those mentioned above is suggested, namely: that proofs are soundly and completely expressed by normal derivations alone (thus, a strengthening of Prawitz's aforementioned thesis, which is acknowledged by Prawitz 1971 himself to follow from his argumentation (see p.258, 4.1.2.1)).

3. Second relevant point of analogy: theses, not conjectures

a. No mathematical statements, therefore no candidates to be mathematically proved or disproved

The cogency of Došen's point that a formal proof of the normalisation thesis should not be expected has already been advocated here (actually, we were, it seems, rather more radical than Došen on this point: he seems to side with Kreisel and Barendregt in stating that the Post-completeness argument is an efficient way of justifying the completeness part of the normalisation thesis once its soundness is granted; a point specifically against which we argued in section III.2.b). The Church-Turing thesis[39], likewise, is usually not treated as subject to formal justification[40]. Turing himself, for instance, in the 1936 article where the thesis as put forward by himself can be first identified, expresses a similar opinion in a most positive way: "All arguments which can be given [to show the thesis] are bound to be, fundamentally, appeals to intuition, and for this reason rather unsatisfactory mathematically". When commenting the closely related 1954 thesis on puzzles, he puts the point in an even more eloquent way: " The statement is moreover one which one does not attempt to prove.

39 Most comments to be made will also apply to the other mentioned related theses.
40 Although there are exceptions: see, e.g. Kripke 2013, who claims that the Church-Turing thesis is a corollary of Gödel's completeness theorem for first-order predicate logic with identity.

Propaganda is more appropriate to it than proof(...)". The reason for such a diagnosis – or, seen from another angle, the diagnosis to which this observation gives rise – can be regarded as the same in all cases: the theses in question, no matter how one employs or understand them, are of a rather *informal* nature – i.e. despite being formulated (at least partially) in mathematical terms, they are no *mathematical* statements (in the sense that they are no statements *of mathematics* as e.g. 1+1=2 is); therefore, they cannot possibly be expected to be decided (or shown undecidable) mathematically. This is, by the way, the first reason why the normalisation thesis – just as much as the Church-Turing thesis – cannot be reasonably regarded as a mere *mathematical* fancy: it is not, strictly speaking, entirely or even essentially mathematical in nature. It is yet to be argued, though, why, if at all, it should not be regarded as a mathematically formulated *metaphysical* or *ideological* fancy, in mathematics or elsewhere.

The terminology "conjecture", most frequently employed when referring to the normalisation thesis, has also been purposefully eschewed here for this reason (though not only). In contexts that are mathematical enough – such as that of the present discussion of identity of proofs –, a conjecture is usually something that is, at least in principle, a candidate to be mathematically proved or disproved – or, at the very least, being mathematically shown to be undecidable.

b. Definitions, not "scientific" conjectures viz. hypotheses: definitions can be used in many ways

Furthermore – and this is the second reason why the label "conjecture" is rejected here – , these theses need not in principle be taken as formal *reconstructions* of some given informal notion. Another aspect they share pointed at by Došen is their *definitional* character; and definitions may well be taken as stipulative, creative, or, more interestingly, *propositive* rather than descriptive in nature. Church, for instance, in the 1936 paper where the thesis is first put forward by himself in terms of λ-definability, refers to his proposal as a "definition of effective calculability"; and rather in the same spirit, one could regard the normalisation thesis as a definition of, say, semantical equivalence between natural deduction derivations. Such definitions could in turn be taken to be attempts to actively, positively *build* or *give substance* to a notion which was previously not there, or that, at best, was too fuzzy, volatile or obscure to serve certain purposes, instead of to truly, faithfully or precisely capture or depict some informal notion assumed to be somehow previously viz. independently determined. Thus, inasmuch as they are taken as attempts at formally reconstructing an informal notion given beforehand, both theses are just as “true” as they are faithful renderings of the given intended informal notion into formally tractable terms. Otherwise, inasmuch as they are proposals as to how we can positively specify and deal with certain ideas not assumed to be previously or independently given or determined in any particular way, these theses should not be looked upon as having any pretension of being “true” in any sense, and are just as good as they are successful, when used as definitions, in allowing viz. blocking uses of the defined notion in accordance with the goals one has, whatever these are. Turing summarises the points made here in V.3.a and V.3.b in the already transcribed comments to the mentioned analogous 1954[41] thesis on puzzles:

“The statement is moreover one which one does not attempt to prove. Propaganda is more appropriate to it than proof, for its status is something between a theorem and a definition. In so far as we know a priori what is a puzzle and what is not, the statement is a theorem. In so far as we do not know what puzzles are, the statement is a definition which tells us something about what they are.”

Where Turing speaks of puzzles in this passage, one could well speak of effective computations or identity of proofs – or even simply proofs –, and the observation would still hold. Thus, in yet another sense, it is a *thesis* and not a conjecture or a hypothesis the proper object of our interest when addressing the normalisation thesis on identity of proofs: the conjecture that can be made by suggesting this thesis as a possibly true hypothesis – of mathematical or other nature – is merely one among the possible ways one can employ the thesis, and it has no privilege in this investigation. Analogously, we do not speak of, say, the normalisation *stipulation* or *postulate* on identity of proofs – for, just the same, a, say, non-veritative way of putting the thesis has no reason to be privileged here. The terminology *thesis*, unsatisfactory as it may be in the discussed respect, follows that which is already sedimented by use in the case of Church-Turing thesis, and seems at any rate less problematic than the considered alternatives.

41 Turing, A.M., 1954, “Solvable and Unsolvable Problems”, Science News, 31: 7–23; reprinted in Copeland 2004b: 582–595

4. Third relevant point of analogy: initial acceptance of soundness, not of completeness

Another important aspect in which Church-Turing thesis and the normalisation thesis are similar is the fact that while their proponents deem their soundness to be somehow obvious, they dedicate significant efforts to argue for their completeness, which they do not take for granted.

A reconstruction of some assumptions under which the soundness part of the normalisation thesis could be taken for granted was given in III.2.a, and some relevant objections to it have also been explored in III.2.a.1 and III.2.a.2. Objections to – or, as a matter of fact, discussion of a satisfactory level of philosophical depth in general of – the soundness of Church-Turing thesis are nevertheless relatively less frequent[42], the standard, overwhelmingly hegemonic attitude being its tacit acceptance.

On the other hand, the completeness of both theses is more frequently questioned, and quite understandably so for especially one before all – many – other reasons: they impose clear and sharp limits to notions manipulated in a great variety of senses which are frequently not sharp at all. This is even more true in the case of proofs and their identity, due to the frequent presence of talk of proofs in most variegated contexts in everyday life. In spite of the fact that the notion of (effective) computability has close relations with distinct notions such as constructibility, feasibility etc., it is usually employed in much more restrict contexts – usually philosophical or mathematical, being quite infrequent in non-theoretical or academic contexts.

5. A brief comparative evaluation of the success of the two theses

a. Church-Turing thesis's unequivocal triumph as the fertile soil of a science, not as one of its fruits

Church-Turing thesis is possibly one of the most successful propositions of its kind. The diagnosis made by Turing as early as 1948 is the following:

42 Two to appear in the literature of which I am aware can be respectively found in Péter, R. *Rekursivität und Konstruktivität*, Constructivity in Mathematics, Amsterdam 1959, pp. 226-233.; and Porte, J. *Quelques pseudo-paradoxes de la "calculabilité effective"*, Actes du 2me Congrès International de Cybernetique, Namur, Belgium,1960, pp. 332-334.

“It is found in practice that L.C.M.s can do anything that could be described as ‘rule of thumb’ or ‘purely mechanical’ [i.e. effectivelly computable]. This is sufficiently well established that it is now agreed amongst logicians that ‘calculable by means of an L.C.M. [a Turing machine]’ is the correct accurate rendering of such phrases.”

Copeland 2017[43], for instance, recognises the virtual consensus amongst logicians on the matter as applicable even to today’s situation – which shows that the thesis is way past the possibility of being called a mere fancy. As Turing exemplifies with his opinion, the reason for such high acceptance is usually attributed to “practice”: In the first place, in tune with Turing’s own way of putting the matter, all examples of functions deemed effectively computable that have been considered so far actually happen to be computable by Turing machines. But perhaps even more importantly, all relevant concurrent attempts to give mathematically tractable substance to the notion of effective computability – in particular Church’s and Turing’s – have turned out to be extensionally equivalent in spite of the sometimes very distinct ideas and formal apparatus in terms of which they are formulated. This is actually sometimes described as indicative of a *formalism independent* character of the thesis’s claim: the exact same class of functions is, after all, selected as effectively computable in all approaches.

But notice that practice only offers “arguments” – and quite indecisive ones – to support, first, the completeness part of the thesis: for the lack of knowledge of counterexamples – or of how they could be produced from potential examples – and the overwhelming variety of examples show, if anything, our lack of concrete material to believe that the thesis may leave out something that should be regarded as an effective computation; and second, the non-ideological character of the very notion of effective computability yielded by the mentioned theses: the extensional coincidence of all considered ideologies arguably suggests this much. But what about soundness? Which aspect of practice offers any arguments to favour the idea that everything that Turing machines can compute is in fact effectively computable or purely mechanical?

I believe that a proper look at why people have been persuaded of Church-Turing thesis’s soundness in the past and nowadays is perhaps one of the best ways to understand

43 Copeland, B. Jack, "The Church-Turing Thesis", The Stanford Encyclopedia of Philosophy (Winter 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2017/entries/church-turing/>.

both the reason and the meaning of the hegemonic acceptance of the whole thesis. It is fairly clear that the thesis was put forward – at least in the versions proposed by Church and Turing – as an attempt not so much to *reconstruct*, but rather to make a certain informal, in some senses deemed “vague” notion of effective computability or calculability a *precise* one. Now, such an enterprise involves establishing sharp limits for the notion to be defined; but this task obviously cannot be guided simply by rigorous observance of the limits of the aimed notion itself. In fact, there was no such thing as limits that *could* be rigorously observed attached to the informal notion in question. In this sense, rather than a reconstruction, Church-Turing’s thesis is, not only from a historical but also from a conceptual, semantical viewpoint, more likely a *contribution*; an actual *addition* to a previous notion of computability which was in fact *transformed* by it. The evaluation of the soundness of the thesis, thus, cannot really be understood as mere comparison with some related previous informal notion of computability which is supposedly to be successfully mimicked by it; it is simply peremptory to even assume there was such thing.

This suggestion goes against, it seems, Copeland’s 2017 view on the matter. He claims that “effective”, “mechanical”, “systematic” etc. are informal “terms of art” in logic, mathematics and computer science that depart significantly from their everyday usage in these contexts. He goes on, and states that these terms are all synonymous when employed in these specific contexts, and that their meaning is as precisely determined as described by the following clauses, which I quote:

“A method, or procedure, M, for achieving some desired result is called ‘effective’ (or ‘systematic’ or ‘mechanical’) just in case:

1.M is set out in terms of a finite number of exact instructions (each instruction being expressed by means of a finite number of symbols);
2.M will, if carried out without error, produce the desired result in a finite number of steps;
3.M can (in practice or in principle) be carried out by a human being unaided by any machinery except paper and pencil;
4.M demands no insight, intuition, or ingenuity, on the part of the human being carrying out the method.”

Now, as much as this suggestion has a reasonable claim to hold with respect to today’s employment of this vocabulary in the mentioned fields – which, more than mere manipulation of terminology “of art”, is rather *technical* –, it is hard to see evidence that the mathematical community uniformly employed – let alone was *aware* of employing – an informal notion of effectiveness as clearly determined as the one just depicted *before* the publication of, more than any other, Turing’s 1936 own groundbreaking *informal* account of the notion. The fact that reasonable – yet thus far clearly defeated from a historical viewpoint – criticism from quite capable practitioners of mathematics has arisen until at least as late as 1960, as already observed, specifically against the soundness of the thesis speaks in favour of this.

Moreover, the notion of computability by a Turing machine, besides being a remarkably accurate formalisation of Turing’s own informal understanding of computability – which is the actual soul of his contribution to the debate on effective computability –, allows for a mathematically very fruitful way of employing the until then somewhat elusive notion of effective computability. This last observation could hardly be overstated – we are talking of a formal development which yielded solutions to the halting problem and to the *Entscheidungsproblem* just to start with.

Thus, the fact that the Church-Turing thesis is a remarkably faithful formalisation of Turing’s own proposed informal account of effective computability is hardly the most decisive factor for the initial endorsement of the soundness of this thesis – at least insofar as it is considered as formulated here, namely: “All and only *effective computations* can be carried out by a Turing machine viz. are λ-definable” – ; rather, such endorsement seems to stem from the fact that Turing’s solid informal conception of effective computability was probably as appealing as possible when measured within the mainstream ideological framework on effectiveness and computability at his time. A remarkable illustration of this point is the consideration of Gödel’s appreciation of the significance of Turing’s work for the proper support of the Church-Turing thesis. He remained clearly unconvinced of the thesis even after it was demonstrated by Church and Kleene in 1935 that λ-definability and general recursiveness are extensionally equivalent, only to enthusiastically endorse the thesis after Turing’s account of computability in his 1936 article. Regarding this, he said:
“Turing's work gives an analysis of the concept of "mechanical procedure" (alias "algorithm" or "computation procedure" or "finite combinatorial procedure"). This concept is *shown* to be equivalent with that of a "Turing machine."” (quoted from Davis 1965, p. 72. My emphasis.)

The "concept" referred by Gödel as yielded by Turing's analysis is *not* the formal one of a Turing machine, or else his claim to its being *shown* to be equivalent to that of a Turing machine would be pointless. He further claims that:

"We had not perceived the sharp concept of mechanical procedures sharply before Turing, who brought us to the right perspective." (Quoted in Wang 1974, p. 85)

As observed above, then, the crucial aspect of Turing's account of computability is his convincing philosophical presentation of (a) why a certain informal notion of effective computability is very appealing for clear reasons and (b) how it can be very faithfully formalised in a certain remarkably simple and interesting, fruitful way (i.e. Turing machines); which only as a bonus happens to be extensionally equivalent to other suggested formal criteria to define effective computations. These other formal criteria were never sufficiently underpinned from a philosophical point of view; they were not *shown* to germinate from a previous appealing informal account of effective computability, and thus remained unjustified in their general claims both to soundness and to completeness. Gödel's appreciation of the matter is thus, in this respect, very close to the one presented here.

However, according to the present appreciation of the matter – which in this respect departs from that of Gödel, as will be clear –, the appeal of Turing's informal account of computability resides in the fact that it is both (a) compatible with the vague notion of effective computability employed at the time, in the sense that it is neither *clearly unsound* nor *clearly incomplete* with respect to it; and (b) by extending and transforming this previous vague notion of computability into a more detailed and solid – yet still informal – one, it offers the basis for a mathematical handling and further development of this notion of unparalleled power. This is in fact why one should not take Turing's wording as mere verbal frolic when he talks of *propaganda*, rather than proof, being more adequate to support the thesis: this is true in senses probably deeper than the mathematical one. As just suggested, thus, the decisive factor in the acceptance of the thesis – and most especially of its soundness – has to do with the conceptual and mathematical possibilities it was capable of inaugurating; it is therefore not a matter of accepting it because it depicts faithfully a notion of effective computation which was previously there – somehow "in the air"[44], whatever this means –, but rather because it

44 See e.g. p.51 of Gandy, R., 1988, "The Confluence of Ideas in 1936", in R. Herken (ed.), 1988, The Universal Turing Machine: A Half-Century Survey, Oxford: Oxford University Press: 51–102.

outlines a notion of effective computation which is remarkably *powerful* from both a conceptual and a technical point of view. And it is hardly relevant here whether or not people *believe* they have accepted Church-Turing thesis due to its faithful correspondence to a previous notion of computability; the fact is that such a notion had never really been expressed before Turing, which makes it take a leap of faith to believe that it was even already there beforehand viz. independently of his formulation. This sort of understanding seems to be emblematically exemplified by, again, the attitude of Gödel towards the Church-Turing thesis, who, in the second of his quoted passages above, refers to what seems to be an assumed previously given "sharp" concept to which Turing's analysis corresponds. It is especially in this respect that his view departs from ours. Closer to our viewpoint is the one suggested by Church 1937. Post (Post 1936, p.105) criticised Church's understanding of his own identification of λ-definability and "effective calculability" as a *definition* of effective calculability as masking the fact that it is rather a "working hypothesis", in need of "continual verification"; in reviewing Post's article, Church then responded to the criticism by claiming something close, yet somewhat less radical, to what is stated here about there being no reasons to assume that there is a previous notion of calculability against which the "hypothesis" is to be verified, and that it thus *establishes,* rather than merely describes, the addressed notion:

"[Post] takes this identification as a "working hypothesis" in need of continual verification. To this the reviewer would object that effectiveness in the ordinary (not explicitly defined) sense has not been given an exact definition, and hence the working hypothesis in question has not an exact meaning. To define effectiveness as computability by an arbitrary machine subject to the restrictions of finiteness would seem to be an adequate representation of the ordinary notion, and if this is done the need for a working hypothesis disappears."

An analogy might be useful to summarise the interpretation advocated here. If we think of certain political leaders and ideologies which, coincidentally or not – most probably *not* –, have emerged and gained massive endorsement in the 1930s' West, the massive endorsement of the Church-Turing thesis itself might be better understood. More specifically, this is meant in the following sense: one would have a hard time defending the claim that e.g. national-socialism or Adolf Hitler found massive approval amongst Germans in this decade due to the fact that they merely mirrored viz. efficiently applied a somehow previously constituted body of political convictions these people held beforehand. Rather than mere faithful codification and executors of a previously established popular will, it seems that they achieved the rather more spectacular deed of offering these people a *new* ideology; one that was capable not only of accommodating their old, previous ideals, but also, in adapting, changing and adding most significantly to them, to provide these people with a solid sense of *purpose*. They brought about and embodied what appeared like *clear-cut*, *unambiguous* political guidelines and goals which Germans, due to a series of social and historical circumstances, not only lacked but also longed for, and also, on the basis of their previous, vague, disempowered and ineffectual political convictions and attitude, believed they were *able* and *willing* to endorse. The suggestion here is to understand the triumph of the Church-Turing thesis as sharing a similar structure: rather than a faithful expression of pre-existent ideology, it is a brand new ideology, capable of imprinting to the old one the sense, direction and acceleration it lacked and longed for in a variety of senses. Again: Turing's use of the word "propaganda" is most suggestive if not confined from the outset into the straightjacket of a mere touch of witty humour – which, by the way, it quite possibly was intended to be. In art as in life; except for the fact that national-socialism was rather heavily countered and defeated, while theoretical and practical work involving computation and complexity flourished and triumphed upon the basis of Church-Turing thesis – at least for the time being.

In any case, conceding as much as we can, the Church-Turing thesis is at the very least, if you will, a case where the so-called paradox of analysis finds an instance which clearly goes far beyond a mere theoretical fuss: even if we do accept that it is, as so frequently suggested, the formal version of a mere *analysis* of an independently and previously given notion of effective computability, it is difficult to see how such an analysis could have brought that much transformation in the use of the analysed notion without having a meaning which is essentially different from that of the latter. But then how, in which sense could one still call it a mere *analysis*?

The "propaganda" in favour of the employment of the Church-Turing thesis to guide the handling of notions such as "effective computability" and equivalents has indeed been so successful that a remarkable phenomenon can be observed in the present stage of their discussion: computable by Turing machines, λ-definable and equivalent notions – especially the first of these, due to the very simple and familiar terms in which it is formulated – became indeed *paradigmatic* cases of what one deems effectively computable in an “intuitive”[45] sense. So, in practice, the soundness of the Church-Turing thesis is not altogether detachable from the frequently supposed “intuitive”, “pre-theoretical” departure point of the discussion of computability. At the very least not anymore. In a sense, thus, to deny the soundness of Church-Turing thesis is literally close to nonsensical – which shows that this thesis has succeeded in *establishing the paradigm* of what it is to be effectively computable, rather than in truly describing this notion.

b. The normalisation thesis: a footnote written in pencil

Matters stand quite differently when it comes to the normalisation thesis, as recalling some of the points already made along this work should now make clear. Although explicit and mathematically oriented attempts to account for identity of proofs in a systematic fashion is a rather recent business – even more than that on computability –, we have argued here that discussion of identity of proofs of a rather formal nature can be found in classical philosophical literature since at least as early as the 18th century; and, maybe even more importantly, the core informal concept at stake in the thesis – namely, proof – has been employed as a “term of art” in philosophy, logic and mathematics for almost as long as there have been such disciplines. This means that the philosophical consideration of the concepts addressed by the normalisation thesis go way farther back in history than the relatively young discussion of computability, which one could take pains to trace back to, at the earliest, say, Leibniz and other 17th century thinkers, and which in any case did not acquire some of the most essential features of its shape before Hilbert and his foundational enterprise.

Quite contrarily to the expectations that such a picture could generate, however, the notion of proof at stake in the normalisation thesis has far more clear-cut determined traces than the notion of effectiveness addressed by the Church-Turing thesis: from the very outset, all eventual proofs that cannot be expressed as natural deduction derivations are not contemplated, while in principle all procedures/computations are possible candidates to effectiveness. The normalisation thesis thus works only in an environment where a clear and specific *application* criterion for proofs is granted, and is either evidently false or simply ill-

45 My experience is that the use of this notion – “intuitive” –, while hardly being of any significant utility at all in philosophical discussion, usually comes at the expensive cost of bringing about a great deal of confusion to it. Therefore it is purposefully avoided here as much as possible.

formulated in its absence – unlike what happens in the case of the Church-Turing thesis, which itself provides such a clear application criterion to a notion of effective computation that is not necessarily previously specified in any way.

But before we proceed into our evaluation of the normalisation thesis and how it compares to that of the Church-Turing thesis, maybe now is the proper time for a brief interlude, in which we are to ask ourselves this: What does this mean: a term of art? Where does such a use of a term lie between a technical, i.e. explicitly stipulated in a fix and possibly arbitrary fashion, and a non-theoretical, spontaneous or unreflective one? In the cases that now interest us, as in many others, it is most certainly *not* the case that philosophers, mathematicians and scientists just chose to use a given term to express a certain specific concept of their disciplines out of no particular reason and in a way independent from its meaning in other contexts where its rules of employment are well enough sedimented by linguistic practices. Terms of art become terms of art usually *due* to their behaviour elsewhere in language, and not independently of it, let alone in spite of it. Thus, e.g. "effectiveness" in logic, mathematics and computer science is not some extraterrestrial appropriation of a term which is employed in most other contexts to refer to some distinct, "usual" notion of effectiveness; an appropriation alien to the sense in which one says that e.g. using a blunt razor blade is not very effective for shaving one's beard. Of course one must recognise that e.g. while writing truth tables is a mathematically effective procedure to test whether or not any given sentence of classical propositional logic is tautological, it could hardly be regarded as effective in a practical sense for testing the tautologousness of sentences with many different variables – it may simply take too long, to the point of unfeasibility. But does this mean that we are dealing with different notions of effectiveness in each of these cases? Some fallacies, for instance, may be deemed rhetorically very effective arguments, despite their being clearly epistemically ineffective. Would it not rather be the case of effectiveness – a single, non-technical and democratically accessible notion, which means, quite obviously, the character of methods or procedures that produce or are capable of producing a certain effect – being measured by different standards, according to different requirements and goals involved in respective tasks or activities? This reflection applies to both cases considered here, of course. This means that proofs, as well as their identity, are also not to be looked upon in this investigation as completely isolated instances of notions which are not but homonymous to some supposed everyday notions of proof and their identity; insulated notions, over the determination of the meaning of which logicians, mathematicians and computer scientists would have complete and exclusive power. I fail to see reasons why anyone should feel entitled to consider what would be such linguistic aberrations as relevant, even as mere possibilities, to the present discussion.

Now, turning back to the matter of the normalisation thesis: the points observed above actually highlight one of the most important discrepant aspects between the two compared theses. Very generally stated, it is this: the normalisation thesis and the Church-Turing thesis work as answers to fundamentally different kinds of questions regarding their respective objects. As stated previously: while the Church-Turing thesis is an attempt at providing the notion of effective computation with an *application* criterion, the normalisation thesis, as remarked before, works only under *the previous assumption* of a specific application criterion for proofs, and provides an *identity* criterion for proofs. Thus, while the Church-Turing thesis answers to the question “what is it for a computation to be effective?” – first, in the intensional sense of “what are the necessary and sufficient conditions for the predicate ‘effective computation’ to apply to something?”, and then consequently also in the extensional sense of “to what does the predicate “effective computation” applies?” –, the normalisation thesis needs to first *take for granted* that the answer to the question “what is it for something to be proof?” (asked in the same sense as in the case of effective computation) is “to be expressible as a derivation.”, to only then present itself as an answer to the question “what is *a* proof?” – in the sense of “what are the necessary and sufficient conditions for something that is a proof to be *the proof it is and no other, regardless of which proof it is*?”. Notice that the Church-Turing thesis has absolutely nothing at all to say about the identity or individuation criteria of that which it selects as effective computations.

When Aristotle considers the matter of definitions and essences, for instance, the difference between these two sorts of questions was not exactly an issue. The instances, so to speak, of the essences considered were particular things *in reality*, thus having their individuation and identity trivially guaranteed from the very outset: a distinct instance means a distinct thing, for a thing “*qua thing”* is obviously only identical to itself. Thus, e.g. *rational animal* is a perfectly sufficient definition of man, despite the fact that it gives one no instructions as to how a man is to be distinguished amongst men; i.e. it expresses the *τὸ τί ἦν εἶναι* (“what-it-is-to-be”) of the species man quite perfectly, because the question concerning individuation and identity criteria never really emerges due to the nature of the instances considered. In the setting of the theses we are now considering, however, our reference to computations and proofs as “real things” whose identity is trivially guaranteed and intrinsically non-problematic (if we do refer to such things at all, which seems to be, as remarked elsewhere, assumed by e.g. the most paradigmatic formulation of the normalisation thesis) is always *mediated*, so to speak, by some kind of *linguistic representative*, the identity value of which (if conceived as determined necessarily and sufficiently by what it represents, which again seems to be assumed by the most paradigmatic formulation of the normalisation thesis) is *not* trivially guaranteed in any manner. Thus, the expression of what would be a rough counterpart of the *τὸ τί ἦν εἶναι* of the addressed notions in a contemporary lexicon demands that we make the identity – or, to express things better, the *equivalence* – criterion of such representatives explicit; a task which is demanded by the "critical" approach inherited from the tradition in which the contemporary treatment of these matters is historically inserted. This means that the normalisation thesis *does*, together with the mentioned assumption on which it depends, offer an account of what it is to be *a* proof, while the Church-Turing thesis does *not* – or at the very least not necessarily – offer an account of what it is to be *a* computation; if so, then only a partial one.

The fact that the implications of the normalisation thesis are, in the sense just described, clearly farther-reaching with respect to the depiction of the contours of its object – and by that now we mean *proofs* rather than identity of proofs – than that of the Church-Turing thesis are with respect to its respective object – namely effective computations – could be regarded as a relevant point in the explanation of why the first is so clearly less accepted than the second, even if one only considers their respective claims to soundness. But probably the most drastic factor is really the lack of appeal and potential for publicity of the first within the current ideological *status quo* of the mathematical and philosophical establishment, as opposed to that of the second. While it feels, as already mentioned, almost absurd to contradict the soundness of the Church-Turing thesis in any discussion of its subject-matter that is not really radically deviant from mainstream, it is simply *too easy* to think of quite unextravagant counterexamples to the soundness of the normalisation thesis – some of which were already presented and discussed here. And this is, again, of course not due to one being true and the other false, but rather to the acceptance of one having served, from the very beginning, as a remarkably good springboard for the further development of what was, until its appearance, mainstream ideology on matters to which effectiveness of computations is a central issue, such as e.g. solvability of mathematical problems; while the acceptance of the other seems to allow no clearly foreseeable significant further development towards a clearly more powerful or somehow advantageous way of dealing with any interesting question which is already stated. In fact, it might be relevant in explaining this to consider the reason why identity of proofs as a problem itself, unlike that of effective computability, never really caught the eye of mainstream mathematics: it has not yet been clearly connected with any “bigger” issue in this area. Those who have given it their attention for not exclusively ludic reasons seem to see it as a question worth being answered mainly for what they consider to be its own intrinsic conceptual import, intimately tied to a *philosophical* endeavour, namely getting to understand what a proof is. So, while the Church-Turing thesis was indelibly written on a page in the history of thought which was, until it came about, filled with many questions and practically no answers – the one on effective computability –, the normalisation thesis is still no more than a sketchy draft written in pencil on the page on proofs; a page that has been marked by countless indelible questions and answers since the very dawn of Philosophy. It now disputes a space in this page’s footer – to which attention has been led back as soon as interest in directly and formally investigating proofs was once again awoken –, side by side with less visible yet very interesting considerations, such as those of Kant on identity of proofs, made in the times when formal logic was still virtually exhausted by syllogistic.

6. On a possible interdependence between the two theses

One could be tempted to say that, in this context, the normalisation thesis can work as a partial answer to the more general question concerning the identity of effective computations. Mimicking *mutatis mutandi* the formulation of the question which the normalisation thesis proposes to answer, we could formulate this one so: when do two Turing machines / λ-terms represent the same computation? The idea would thus be that the mentioned background assumption of the normalisation thesis can actually be looked upon as the statement of a particular case of the Church-Turing thesis. As just stated, this assumption consists in the acceptance of a specific *application criterion for proofs* – i.e. roughly, a criterion that determines the necessary and sufficient conditions to be satisfied by something so that the predicate “proof” applies to it. Now, one of the theses mentioned and described previously as more closely analogous to the Church-Turing Thesis than the normalisation thesis itself offers precisely such a criterion – namely, the one claimed by Prawitz 1971, according to which every proof in first order logic can be carried out by means of some Gentzen-style natural deduction derivation, and by means of every such derivation can some such proof be carried out. Thus, providing an identity criterion for proofs could be taken as ultimately amounting to providing an identity criterion for some specific subclass of effective computations. This consideration involves, however, a very significant mistake. More specifically, it would be conditioned by a failure to observe the fact that the notions of proof (the informal one, formally specified by the background assumption of the normalisation thesis, of course) and effective computation (also the informal one, as characterised by Church-Turing thesis), despite having intrinsically related *application* criteria, are still at least in principle essentially distinct viz. not necessarily related – in particular, there is no reason at all to see proofs as a subclass of effective computations, since the identity criteria of the latter are left undefined and might thus be completely distinct from those of the first (as e.g. in the case of positive rational numbers and ordered pairs of positive natural numbers, mentioned in other section of this work). It is not difficult to understand that the background assumption of the normalisation thesis implies *not* that first order logic proofs are all expressible by effective computations of a special class; but rather that these proofs are all expressible by *that which, according to the Church-Turing thesis, also expresses* certain effective computations – namely those formal expressions of effective computations (typed λ-terms, etc.) that somehow correspond (Curry-Howard) to formal expressions of proofs (natural deduction derivations). Nevertheless, it would suffice to assume that first-order logic proofs indeed *are* a particular subclass of effective computations in Turing's informal sense – which would not seem extravagant at all, given the fact that the acceptance of the background assumption of the normalisation thesis is far greater than that of the normalisation thesis itself – to restore the claims just shown to be in principle unfounded.

VI. Concluding remarks: What can we (not) accomplish with the normalisation thesis

The observations made thus far suggest that, even before the normalisation thesis itself, the very question concerning identity of proofs – even if restricted to the version to which the normalisation thesis actually presents an answer –, unlike the one on effectiveness of computation, answered by the Church-Turing thesis, has no clear potential, import or utility to *broader* mathematical and conceptual problems. Its only clear intrinsic connection seems to be to the philosophical question which motivates it, namely: what is a proof?

Answering this particular philosophical question does not seem to be within the prioritised scope of interests of significantly many significant mathematicians or philosophers – in many cases, not even of those few who occupy themselves with proof theoretic investigations. Notwithstanding, it seems that enough propaganda, both in quantity and in quality, could change this situation. History indicates that the piece of publicity I offer here will most probably result completely ineffectual for this purpose. It is a clear fact that the eminent, unavoidable, intrinsic *political* import of proofs in general – which, as suggested in the prologue, has as its provenance and its effect the eminent, unavoidable, intrinsic political import of *knowledge* – has been systematically made invisible in our tradition; a historical move which was crucial so that these conceptual devices could have a chance to *in practice* exert and achieve – in very particular and questionable ways, of course – the political roles and goals Plato devised for them from the very beginning. So, insistently underlining these features as I make here will probably, instead of casting light upon the great importance and interest of improving one's understanding of proofs and their meaning, only make my discourse sound as either rather confused and somewhat hallucinatory or just off the point. Be that as it may, what my lack of talent as an adman leaves me as alternative to my current attitude is either to silence, or to adopt the clearly even less productive option of surrendering to the flow of propaganda that I here counter and, at the very best, adding some more baroquish scholastic meanders to one or more of the still unfinished yet already overburdened epistemic-oriented cathedrals of theory and terminology on proofs, which have so far notoriously failed to illuminate the relevant aspects of this notion in which we are interested in this investigation.

In any case, let us suppose that one could triumph in this enterprise of publicity and bring proofs and their identity to the very center of mathematicians and philosophers considerations and concerns, rather like it was with effectiveness and computability in the first half of the 20th century. What would then be our reasons to endorse something such as the normalisation thesis?

It is easy to see that any reasons to do this much would not be “scientific”: no informal notion of proof that is not too fabricated and/or partial to perform the (already chimerical enough) task of “capturing what is essential about proof” corresponds faithfully to that formally outlined by the normalisation thesis. Those that do, e.g. a proper version of BHK, are interesting enough to explain certain phenomena about proofs as semantical counterparts of derivations; but obviously its coverage of the addressed notion is not enough. As already shown and stressed, the counterexamples are too many, too significant and too unextravagant.

This brings as a consequence that the normalisation thesis could also not serve as a “scientific” springboard to the investigation of proofs even nearly as efficiently as the Church-Turing thesis served to the investigation of effective computation. Since too much obvious and obviously interesting material cannot be handled as proofs under its acceptance, eventual results obtained would lack the necessary conceptual comprehensiveness to yield acceptable enough characterisations of what a proof is. The background assumption of the normalisation thesis, however, which is, as observed above, significantly more analogous to the Church-Turing thesis than the normalisation thesis itself, has been exerting the role of such a springboard – though, it seems, with much less coverage with respect to proofs (unless, arguably, we were to restrict ourselves to a realm of “effective” first-order proofs) than the Church-Turing thesis has with respect to effective computations. This is of course hardly surprising: for as already mentioned, the historical development and provenance of the notion of proof is significantly more intricate and remote than that of the notion of (effective) computation, which makes it sound rather reasonable that it should be expected to be more difficult to coherently and cohesively describe, let alone formalise.

But propaganda, it seems, could in principle change even this. In such a case, we would of course be talking of an enterprise of publicity far more prodigious than overruling some relevant and sensible yet few voices contrary to the soundness or to the completeness[46] of a certain notion of effective computation. Notwithstanding, the right amount of time and the wrong kind of interests could really allow for portentous possibilities. When considering related issues in the conclusion of his 2003, Kosta Došen says that:

46 For a case built specifically against the compleneteness of the Church-Turing thesis, see Kalmár, L., 1959, “An Argument Against the Plausibility of Church’s Thesis”, in A. Heyting (ed.), 1959, Constructivity in Mathematics, Amsterdam: North-Holland: 72–80.

“The complaint might be voiced that with the Normalization (…) [Conjecture] we are giving very limited answers to the question of identity of proofs. What about identity of proofs in the rest of mathematics, outside logic? Shouldn’t we take into account many other inference rules, and not only those based on logical constants? Perhaps not if the structure of proofs is taken to be purely logical. Perhaps conjectures like the Normalization (…) [Conjecture] are not far from the end of the road.

Faced with two concrete proofs in mathematics—for example, two proofs of the Theorem of Pythagoras, or something more involved—it could seem pretty hopeless to try to decide whether they are identical just armed with the Normalization Conjecture(...). But this hopelessness might just be the hopelessness of formalization. We are overwhelmed not by complicated principles, but by sheer quantity.”

The question would then be: would this be desirable? Do we *want* something as the normalisation thesis to take over as the touchstone of our investigations concerning proofs? If so, why? Given all the points raised previously against the soundness and the completeness of the normalisation thesis – points that purposefully do not even resort to the issue of “limitedness” mentioned by Došen – it seems to me that our manipulation of the notion of proof would inescapably have to be severely distorted and deformed in order to fit into the straightjacket the normalisation thesis offers as a royal garb.

But let us suppose for a moment that I can be proved wrong, and the apparent shortcomings of the normalisation thesis could be somehow circumvented, by flares of wit or sheer diligence. One would then have shown that it is possible to “analyse”, to systematically and generally rewrite our use and talk of proofs, inasmuch as it lies within the scope of the normalisation thesis, in terms of the notion of proof that is determined by this thesis. One could maybe even get to show that only a very special aristocracy of notions of proof is capable of playing such a role, i.e. allowing that our use and talk of proofs is organised to form a neatly structured edifice, the architecture of which, no matter how complex or large in dimensions, rests stolid upon one single point.

There are, however, severe problems with this sort of enterprise. One is that people usually – and quite understandably, though not really justifiably – fear too much that their edifices go down once they put too much efforts into building them. If the foundational points are made of sand, or the edifice walls of playing cards, they will most probably try to turn them into cement and concrete with the alchemy that only propaganda makes possible – they would then not only have become indestructible, but also impossible to reorganise or restructure without employing a great deal of demolishing power and getting rid of large, heavy masses of debris and rubble. Another is that people also tend to lose interest in other possible ways of looking at a certain phenomenon once they succeed in organising it in as systematic and exhaustive a way as they can – especially if a great deal of effort or passion was involved in this. I believe – and for reasons that I believe to be clear enough – that these frequent phenomena are notoriously deleterious to the enterprise of *understanding* matters that we address; which seems to me to be something of the utmost importance in philosophical endeavours.

Therefore, I tend to see no sufficiently good reason why investing in such possibilities of systematisation of the semantical and philosophical account of proofs as the one the normalisation thesis seems, at least in principle, to offer. They are *per se,* it seems, not at all pernicious – but it just happens that, within history, nothing is ever *per se*. And it is history that shows their potential and tendency to lead to perfectly justified ideological totalitarianisms, which I most understandably fear. Alas, as Wittgenstein noticed already in those days, most still tend to proudly wallow in craving for the generality I so frequently dread.

VII. Bibliographic References

M. Aigner and G. M. Ziegler. Proofs from THE BOOK, Springer- Verlag, (1998).

T. Aoto, Uniqueness of Normal Proofs in Implicational Intuitionistic Logic, Journal of Logic, Language and Information 8, pp. 217–242, (1999).

Aristotle, Aristotle's Prior and Posterior Analytics, a revised text with introduction and commentary by W.D. Ross, Oxford, (1949).

Aristotle, Aristotelis Topica et Sophistici Elenchi, ed. W.D. Ross, Oxford, (1958).

Aristotle, Hugh Tredennick and E. S. Forster (trans.), Posterior Analytics. Topica. Harvard University Press, Loeb Classical Library 391, (1960).

Aristotle, Aristotelis Categoriae et Liber de Interpretatione, ed. L. Minio-Paluelo, Oxford, (1949). (English translation: Aristotle, Aristotle's Categories and De Interpretatione, translated with notes and glossary by J.L. Ackrill, Clarendon Aristotle Series, Oxford, (1963).)

Aristotle, Prior Analytics, Translated, with introduction, notes, and commentary by Robin Smith. Hackett Publishing Company, Indianapolis,Cambridge (1989).

Aristotle, Aristotle's Prior Analytics book I. Translated with an introduction and commentary by Gisela Striker (Clarendon Aristotle Series) - Oxford University Press (2009).

J. Barnes, Speusippus and Aristotle on homonymy, in Logical Matters: Essays in Ancient Philosophy II, Oxford University Press, Oxford (2012), pp.284 – 311.

R. Carnap, Meaning and Necessity: A Study in Semantics and Modal Logic, The University of Chicago Press, Chicago, (1947).

C. Celluci. Proof theory and theory of meaning, Italian studies in the philosophy of science (M. L. Dalla Chiara, editor), D. Reidel Publishing Company, (1981), pp. 13-29.

A. Church, An Unsolvable Problem of Elementary Number Theory, American Journal of Mathematics, 58 (1936), pp. 345–363.

A. Church, Review of Turing 1936, Journal of Symbolic Logic, 2 (1937a), pp. 42–43.

A. Church, Review of Post 1936, Journal of Symbolic Logic, 2 (1937b), p.43.

A. Church, On Carnap’s analysis of statements of assertion and belief, Analysis, vol. 10 (1950), pp. 97-99.

B. J. Copeland, The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, Oxford: Oxford University Press (2004).

B. J. Copeland, The Church-Turing Thesis, The Stanford Encyclopedia of Philosophy (Winter 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2017/entries/church-turing/>.

M. Davis, Why Gödel Didn't Have Church's Thesis?, Information and Control 54, (1982) pp. 3-24
M. Davis (ed.), The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions, New York: Raven (1965).

H. Diels, and W. Kranz, Die Fragmente der Vorsokratiker, 6th edn. Berlin: Weidmann (1951–52).

K. Došen and Z. Petrić, The maximality of the typed lambda calculus and of cartesian closed categories, Publ. Inst. Math. (N.S.) 68(82) (2000), pp. 1-19 (available at: http:// arXiv. org/ math. CT/ 9911073).

K. Došen and Z. Petrić, The maximality of cartesian categories, Math. Logic Quart. 47 (2001), pp. 137-144 (available at: http:// arXiv. org/ math. CT/ 9911059).

K. Došen and Z. Petrić, Isomorphic formulae in classical propositional logic. Math. Logic Quart. 58(1-2)(2012), pp. 5–17

K. Došen, Identity of proofs based on normalization and generality, The Bulletin of Symbolic Logic, vol. 9 (2003), pp. 477-50

M. Dummett, The logical basis of metaphysics. Duckworth, London (1991).

J. Ekman. Normal proofs in set theory. Ph. D. thesis. University of Göteborg (1994).

S. Feferman, Review of Prawitz 1971. J. Symbolic Logic 40 (1975), pp. 232-234.

G. Frege, Über Sinn und Bedeutung. Zeitschrift für Philosophie und philosophische Kritik, NF 100, (1892) pp.25-50

R. Gandy, The Confluence of Ideas in 1936, in R. Herken (ed.), 1988, The Universal Turing Machine: A Half-Century Survey, Oxford: Oxford University Press (1988), pp.51–102.

G. Gentzen, Untersuchungen über das logische Schließen, Math. Z. 39 (1935), pp. 176-210, 405-431 (English translation: Investigations into logical deduction, in: The Collected Papers of Gerhard Gentzen, M.E. Szabo ed., North-Holland, Amsterdam, 1969, pp. 68-131).

N. Goodman, On Likeness of Meaning, Analysis, 10 (1949) pp.1–7.

S. Hirokawa, The Converse Principal Type-Scheme Theorem in Lambda Calculus, Studia Logica 51 (1992), pp. 83-95.

J. R. Hindley. Basic Simple Type Theory. Cambridge University Press, (1997) C.B. Jay, N.Ghani, The virtues of eta-expansion. J. Functional Programming 5 (2), Cambridge University Press (1995), pp. 135-154

L. Kalmár, An Argument Against the Plausibility of Church's Thesis, in A. Heyting (ed.), Constructivity in Mathematics, Amsterdam: North-Holland, (1959) pp. 72–80.

I. Kant, Die falsche Spitzfindigkeit der vier syllogistischen Figuren, Königsberg (1762).

I. Kant, G. Lehmann (ed.), Vorlesungen über Logik. Teilband 1 (Gesammelte Schriften Bd.24 – 1)-Walter de Gruyter (1966)

I. Kant, G. Lehmann (ed.), Vorlesungen über Logik. Teilband 2 (Gesammelte Schriften Bd.24 – 2)-Walter de Gruyter (1966)

Y. Komori and S. Hirokawa, The number of proofs for a BCK-formula, The Journal of Symbolic Logic, 58 (1993), pp. 626-628.

G. Kreisel. Mathematical logic. Lectures on modern mathematics (T.L. Saaty, editor), vol. III, John Wiley & Sons (1965), pp. 95-195.

G. Kreisel, A survey of proof theory II, in: J.E. Fenstad ed., Proceedings of the Second Scandinavian Logic Symposium, North-Holland, Amsterdam (1971), pp. 109-170.

S.A. Kripke, The Church-Turing 'Thesis' as a Special Corollary of Gödel's Completeness Theorem, in B.J. Copeland, C. Posy, and O. Shagrir (eds.), Computability: Turing, Gödel, Church and Beyond, Cambridge, MA: MIT Press. (2013), pp. 77–104.

Y.N. Moschovakis, What is an algorithm?, in: B. Engquist and W. Schmid eds, Mathematics Unlimited—2001 and Beyond, Springer, Berlin (2001), pp. 919-936.

F. Nietzsche, Die Philosophie im tragischen Zeitalter der Griechen, Entstanden 1872/73. Neuausgabe mit einer Biographie des Autors. Herausgegeben von Karl-Maria Guth. Berlin (2017).

R. Péter, Rekursivität und Konstruktivität,in A. Heyting (ed.), Constructivity in Mathematics, Amsterdam: North-Holland (1959), pp. 226-233

Plato, Platonis Opera, ed. John Burnet. Oxford University Press. (1903)

Plato. Theaetetus, in: Plato in Twelve Volumes, Vol. 12 translated by Harold N. Fowler. Cambridge, MA, Harvard University Press; London, William Heinemann Ltd. (1921).

Plato. Timaeus, in: Plato in Twelve Volumes, Vol. 9 translated by W.R.M. Lamb. Cambridge, MA, Harvard University Press; London, William Heinemann Ltd. (1925).

Plato, Phaedo, in: Plato in Twelve Volumes, Vol. 1 translated by Harold North Fowler; Introduction by W.R.M. Lamb. Cambridge, MA, Harvard University Press; London, William Heinemann Ltd. (1966).

Plato, Meno, in: Plato in Twelve Volumes, Vol. 3 translated by W.R.M. Lamb. Cambridge, MA, Harvard University Press; London, William Heinemann Ltd. (1967)

Plato. Republic. Plato in Twelve Volumes, Vols. 5 & 6 translated by Paul Shorey. Cambridge, MA, Harvard University Press; London, William Heinemann Ltd. (1969).

J. Porte, Quelques pseudo-paradoxes de la “calculabilité effective”, Actes du 2me Congrès International de Cybernetique, Namur, Belgium (1960), pp. 332-334.

E.L. Post, Finite Combinatory Processes – Formulation 1”, Journal of Symbolic Logic, 1 (1936), pp. 103–105.

D. Prawitz. Natural Deduction. A Proof-Theoretical Study. Stockholm (1965), Almqvist & Wiksell

D. Prawitz, Ideas and results in proof theory, in: J.E. Fenstad ed., Proceedings of the Second Scandinavian Logic Symposium, North-Holland, Amsterdam (1971), pp. 235-307.

D. Prawitz, Philosophical aspects of proof theory, in: G. Fløistad ed., Contemporary Philosophy: A New Survey, Vol. 1, Nijhoff, The Hague (1981), pp. 235-277.

D. Prawitz, *Comments on the papers*, Theoria 64 (Special issue on the philosophy of Dag Prawitz), (1998), pp. 283-337.

P. Schroeder-Heister., A Natural Extension of Natural Deduction, The Journal Of Symbolic Logic, Volume 49, Number 4 (1984), pp.1284-1300.

P. Schroeder-Heister., The calculus of higher-level rules, propositional quantification, and the foundational approach to proof-theoretic harmony . Draft. To appear in revised form in Studia Logica, Special issue, ed. A. Indrzejczak, "Gentzen's and Jaśkowski's Heritage: 80 Years of Natural Deduction and Sequent Calculi", (2014).

G. Sundholm, Identity: Absolute. Criterial. Propositional. in: Timothy Childers (ed.), The Logica Yearbook 1998, The Institute of Philosophy, Academy of Sciences of the Czech Republic, Prague, (1999) pp. 20-26.

G. Sundholm, Proofs as Acts versus Proofs as Objects: Some Questions for Dag Prawitz, Theoria 64 (2000), pp. 187-216

G. Sundholm, Semantic Values for Natural Deduction Derivations, Synthese, Vol. 148, No. 3, Proof-Theoretic Semantics (2006), pp. 623-638.

M. Tatsuta. Uniqueness of normal proofs of minimal formulas. The journal of symbolic logic, vol. 58 (1993), pp. 789-799.

A.S. Troelstra, Non-extensional equality, Fund. Math. 82 (1975), pp. 307-322.

A.M. Turing, On Computable Numbers, with an Application to the Entscheidungsproblem", Proceedings of the London Mathematical Society (Series 2), 42 (1936–37), pp. 230–265; reprinted in Copeland 2004, pp. 58–90.

A.M. Turing, Solvable and Unsolvable Problems, Science News, 31, (1954), pp. 7–23; reprinted in Copeland 2004, pp. 582–595

H. Wang, From Mathematics to Philosophy, New York: Humanities Press (1974).

F. Widebäck, Identity of Proofs, doctoral dissertation, University of Stockholm, Almqvist & Wiksell, Stockholm (2001).

L. Wittgenstein, The Blue and Brown Books. Oxford: Basil Blackwell (1958).

L. Wittgenstein, Philosophische Untersuchungen. Kritisch-genetische Edition. Herausgegeben von Joachim Schulte. Wissenschaftliche Buchgesellschaft. Frankfurt (2001).

L. Wittgenstein, Tractatus Logico-Philososophicus; Edição bilíngue alemão-português. Tradução, apresentação e estudos introdutórios de Luiz Henrique Lopes dos Santos; [introdução de Bertand Russell]. – 3ªed. – São Paulo: Edusp (2010)