Philosophica 72 (2003) pp. 105-130

THOUGHT EXPERIMENTS, RHETORIC, AND POSSIBLE WORLDS

Benoît De Baere

ABSTRACT

As early as the seventeenth century, Copernicus' heliocentric cosmology raised the question of a suitable cosmogony. In fact, Descartes' *Principia philosophiae* (1644) and *Le monde* (1633, publ. 1664) marked the beginning of a quite important tradition in Western philosophical history, which spanned the seventeenth, eighteenth and the first half of the nineteenth century. As it turned out, most of those cosmogonies related to one another in polemical ways. However, they shared at least one key feature: their desire to offer some sort of knowledge. But although this "knowledge" had to fit certain observational data, it could not possibly be established on an experimental basis: it could only be thought of, and narrated in a particular way. For that reason cosmogonies were (and still are) basically thought experiments (henceforth: TEs).

To the extent that most cosmogonists cannot bring themselves to present "just" a story, but have the ambition to offer true or at least very probable histories, their narratives need to be studied through the following questions: *What* histories do they tell us? *How* are these stories led to their conclusions? And *why* are these stories told? My aim is to show that the responses to these questions pertain to the fact that philosophers, when they present a cosmogony, posit a TE. I will first demonstrate that the function of those TEs is argumentative and illustrative rather than heuristic: they are *exempla*, i.e. arguments, in larger-scale philosophical demonstrations. As such, they boast two features: (1) they tend to respect the criteria of verisimilitude rather than those of truthfulness, and (2) their make-up is determined by their desired outcome (a world exactly like ours) rather than by their starting conditions. In the second part of this paper, I will explain — by means

of a text-semiotic possible-worlds analysis — that this does not necessarily result in a merely "fictional" discourse.

PART I
THOUGHT EXPERIMENTS, RHETORIC, AND VERISIMILITUDE

1. Thought Experiments and Truth: A Semiotic Approach

Philosophical discourse owes itself a definition of its objects, even if a characterisation that would be able to subsume the diversity of TE's forms and applications remains to be found. And so, instead of accepting Brown's disenchanted suggestion that "TEs are performed in the laboratory of the mind" and that "beyond that bit of metaphor it's hard to say just what they are. We recognize them when we see them" (1991, p. 122), I will turn to Irvine's minimalist account according to which "TEs are to be understood as arguments concerning particular hypothetical events or states of affairs" (1991, p. 158):

> A TE is an instance of reasoning, which attempts to draw a conclusion about how the world is or could be by positing some hypothetical, and perhaps even counterfactual, state of affairs. In short, it is an instance of hypothetical reasoning whose antecedent assumptions may well be false but which leads us to conclusions about the nature of the world or about our surroundings. (*Ibid.*, p. 149)

I am well aware of the fact that this definition, when referred to in such a blunt way, pays no heed to some very important issues. Still, it gives me a starting point we can, I hope, agree on. I will only add, once again with Irvine, that not every instance of hypothetical reasoning constitutes a TE. Indeed, TEs "must stand in a privileged relationship both to past empirical observations and to some reasonably well-developed background theory" (*Ibid.*, p. 150). Moreover, a "TE needs to be acted out in enough detail (in a controlled enough environment, so to speak)" to make it falsifiable (*Ibid.*, p. 159). As for Irvine's condition that a TE "must be relevant to the testing of some hypothesis (or to the answering of some set of questions) which has arisen within a particular observational/theoretical context", I think it is unnecessary: Hofstadter

and Dennett have shown the existence of TEs that, "however systematically developed, are intended merely to illustrate and enliven ideas". Sometimes, indeed, "the boundaries between proof, persuasion, and pedagogy cannot be drawn" (1982, p. 459) — which brings me to my point: a TE is also (and one might even say: *above all*) a "semiotic artefact" (a text) that purports to convey a truth.

Text semiotics considers truth to have different faces. Plett, for instance, distinguishes between the following conceptions of textual truth:

> Ein faktizistischer, ein logischer und ein kommunikativer. Der erste verankert die Bestätigung der Textaussage im bezeichneten Objekt, der zweite in der argumentativen Schlüssigkeit der Zeichenfolge, der dritte in der Kommunikationssituation. Dieser Dreiheit entspricht eine dreifache Semiose des Textes: die semantische (im engeren Sinne), die syntakto-semantische und die pragma-semantische. (1979, p. 101)

This means that by defining TEs through the indirect but nonetheless privileged relations they have with the real world (or rather, with already systematised knowledge about it) Irvine (1991) focuses on what Plett calls "semantic semiosis *stricto sensu*". Norton (1991 and 1996), on the other hand, singles out the "semantico-syntactic" aspect of a TE's semiotic functioning by stating that all legitimate TEs are arguments with premises grounded in experience and that they reach a conclusion based on deductive or inductive inference rules. And finally, when Hofstadter and Dennett (1982) show some TEs to have a merely illustrative scope, they point out the importance of the "semantico-pragmatic" component of TE's semiosis.

Even if an analysis of TEs should ideally take into account all three aspects, I will only focus on the "semantico-pragmatic" semiosis, as I believe it provides a justification for the links I establish between (some) TEs and rhetorical *exempla*.

2. Cosmogonic Narratives as Thought Experiments

2.1 A Cartesian Thought Experiment

Descartes' interest for cosmogonic speculation can be traced back to the

early 1630s: between 1629 and 1633 he composes a treatise entitled *Le monde, ou traité de la lumière* (*The World, or a Treatise on Light*). As it was printed only in 1664, it was never published during his life. Still, it appears Descartes never abandoned his hope to publish this treatise himself: in the fifth chapter of his *Discourse on method* he gives a short but very accurate presentation of it, and in the *Principia philosophiae* he restates its claims, streamlines its narrative, and expands its scope.

Descartes begins his cosmogony by asking his reader to leave the real world behind and to join him in the "imaginary spaces" (*les espaces imaginaires*) where he will witness the birth of a "new world" exactly like ours (1996, vol. VIII p. 31 ff.). Those spaces are empty and as their length, breadth and depth extend well beyond our perceptions, they amount to pure and indefinite extension. As such, they provide the stage where the cosmogonic thought experment can take place.

Subsequently, Descartes introduces — i.e., stipulates — the matter that constitutes his world: the stars, the planets... Assuming that God creates so much matter that there are no empty spaces left, he asks to conceive of this matter in a way that would be perfectly understandable: a solid that fills everything and has no other characteristics but its extension. This extension, now, is of fundamental importance to the rest of his hypothesis. For he goes on, supposing that God divides this matter in particles which he forces to move:

> Now let us suppose that all the difference He establishes [between the various particles of matter], consists in the diversity of the movement He gives to them. As a consequence, as soon as those particles are created, some start to move in one direction, others in the opposite direction; some faster, others slower (or even, if you prefer, not at all). Afterwards, they continue their movements according to the laws of nature. (*Ibid.*, p. 33. My translation)

Descartes believes that these laws of nature (in fact, they are laws of movement) allow for the self-organisation of matter into an orderly universe. In a plenum, he says, a particle can only move to the extent that the ones surrounding it move as well, and take the place it first occupied. Hence, Descartes states, all movement is circular and creates little vortices or "eddies" (*tourbillons* — *Ibid.*, p. 48 ff.):

> All movements in the world are, in one way or another, circular. This means that, when a body leaves its place, another one has to take it, and another one, the place of the second particle, and so on, until the last one: it should occupy the spot left by the first one at the same moment this particle leaves it. Consequently, when particles of matter move, there is no more vacuum between them than when they are at rest. (*Ibid.*, 19. My translation)

Descartes then turns to one eddy in particular. Its movement, he says, rubs particles of matter against each other. This makes them lose their sharpest edges (we can observe this phenomenon when we look at the stones carried by the waters of a river) and gives them new forms: some are divided into smaller particles, others agglutinate; some accelerate, others lose speed.

In the Cartesian system, this amounts to an explanation of the differences between the three elements. For if fire is nothing but a collection of extremely small and fast-moving particles, and if air consists of particles of intermediate size and speed while earth is nothing but a lump of bulky and slow particles, the particles of matter that are blunted can be identified to the second element (air), the scrapings that result of this action, to the first element (fire), and the particles that formed clusters, to the third element (earth)[1].

Still, many more particles of the first element are produced than is needed to fill the spaces between the second and third elements; as a result, they tend to retire to the centre of the eddy where they form perfectly spherical, liquid, and subtle, bodies: the sun and the stars. At the same time, the biggest particles stick together and form planets and comets: dragged along by the particles of the second element they rotate around the centre of the vortex (*Ibid.*, p. 51-52).

At this point Descartes feels he has explained the structure of the solar system, using mechanical terms only. That is why the actual cosmogonic narrative of the *World* ends here, and makes way for considerations on the nature and characteristics of light. However, we need not concern ourselves with these matters: the preceding summary of his story allows me to make my point.

---

[1] It is only in the *Principia* (IV, 139 et IV, 144 sq.) that Descartes introduces a fourth element.

It might not show in my paraphrase, but Descartes places this series of events under the sign of verisimilitude and necessity. This becomes clear when we consider the many occurrences of the words "likely" (*vraisemblable*) and "necessary" (*nécessaire*) in his treatise[2]. At the same time, however, Descartes insists on his cosmogony being only a fiction[3]. As a consequence Descartes' *World* posits a strictly hypothetical (and even counterfactual) state of affairs, which constitutes the starting point of a cogent argument related to a particular "background theory": the idea that since the world is a kind of machine, its operation can and should be explained in strictly mechanical terms.

Even if the Cartesian cosmogony can be seen as a TE, its function is not heuristic. It is only meant to illustrate the efficiency, and consequently to establish the legitimacy of the mechanist paradigm, the *instar machinæ* that gives rise to Cartesian physics. In the same way a functioning automaton "proves" the quality of the diagrams according to which it has been built, the success of the Cartesian cosmogony proves mechanist physics to be a solution to the "cipher" of the world[4]. As such, the narrative of the *Le monde* and its reprise in the *Principia* exemplify the kind of TEs Hofstadter and Dennett showed to "illustrate" and "animate" ideas, whereas Brown called them "mediative" because they "facilitat[e] a conclusion drawn from a specific, well-articulated theory" by showing that "what was unthinkable is not so unthinkable after all" (1991, pp. 36-37).

---

[2] For some cases in point, see *Le monde* (vol. XI of the *Œuvres complètes*) at pages 7, 9, 19, 23, 29, 31, 35, 37 (two occurrences) 44, 52 (two occurrences), 57, 65, 76, 102, 108, 114, 118.

[3] *Cf.* Blumenberg (1999, p. 234 ff.), Hallyn (1999, p. 39 ff.), and De Baere (2004, p. 206 ff.).

[4] *Cf.* Toulmin: "In closing the *Principles of Philosophy*, [...] Descartes refuses to claim logical or metaphysical certainty for his account of nature. He cannot formally prove that his system of natural philosophy is the one-and-only theory free of contradiction or inconsistency. We are to think of it, rather, as one tentative way of deciphering natural phenomena, and, as such, it has only a *moral* certainty. Still, 'moral certainty' is not to be despised. Faced with a script whose sense we do not understand, we are happy to reach a point at which we can interpret its symbols in ways that make sense at all: the more examples an interpretation lets us read without lapsing into unintelligibility, the more confident we are that we have in essentials hit on their actual meaning." (1992, p. 73) For an interesting take on this problem, see Hallyn (2001).

2.2 Criticising the Cartesian Cosmogony: Noël-Antoine Pluche

In his 1739 best-seller *The History of the Heavens* (*Histoire du ciel*), the French clergyman Noël-Antoine Pluche tried — among other things — to show that Descartes' TE is fundamentally flawed. In this paper, I will focus on two aspects of this criticism: his attack on the legitimacy of the geometric determinism (the "logic") that underlies Descartes' cosmogony, and the way he points out precise, but significant, errors.

(1) By refusing to accept the identification of matter and extension, Pluche attacks the very logic on which Descartes' TE is grounded. Indeed, he says, if there were empty spaces between the particles of matter, the geometric (mechanistic) determinism that underlies the *World*'s narrative would be impossible to maintain. As a result, Descartes needs to preclude even the logical possibility of a vacuum — and the only way to achieve this is to identify matter and extension. Pluche, however, argues that Descartes is unable to do this.

> The point on which I shall stop him, is his pretence that a vacuum is impossible: nor is it even so in his own supposition; for in order to fill up all the interstices, there must be dusts of all shapes that come timely and get into the open interstices. These dusts are formed only in a length of time. The globules are not instantaneously made round. The biggest corners are first broken, then the smaller; and, by repeated frictions, we possibly shall collect enow [*sic*] of our pulverized pieces to fill up whatever we shall please. But this pulverisation is successive. The first moment therefore God shall put the particles of the primitive matter in motion, the dust is not yet formed. God makes angles to rise. They will soon begin to be bruised: but, before the thing is done, there are an infinitude of empty spaces between these angles, and nothing is provided to fill them up. (1741, vol. II, pp. 164-165)

This argument is highly significant, because to formulate it Pluche had to think along the lines of Descartes' mechanistic determinism. It is "from the inside", so to speak, that he shows the deficiency of the starting conditions which the *World*'s cosmogonic hypothesis is built on.

(2) Anther component of Pluche's attack on Cartesian cosmogony resides in his insistence on several errors in the *World*'s (and, for that matter, the *Principia*'s) argument. As a matter of fact, Pluche reckons that Descartes did not only posit inadequate starting propositions for his TE, but,

moreover, that the way he develops it is scarcely more satisfying.

He gives numerous examples to substantiate this claim, but the most spectacular one is undoubtedly his demonstration that the shape of Descartes' sun, and indeed his whole world, is not spherical at all. Consider the figures *1* and *2*:

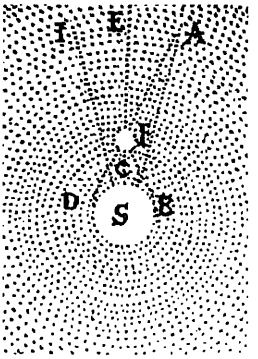

fig. 1

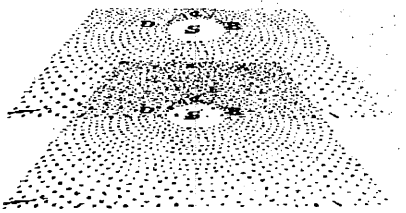

fig. 2

In the *Principia philosophiae*, Descartes goes to great lengths to explain why "the sun and the fixed stars are globes". He formulates the following hypothesis. All the small particles of matter that turn around the centre *S* of the eddy *IEA* (see fig. 1) tend to move away from the centre *S* according to the laws of movement. The presence of the particles of the second element, however, does not allow them to follow the tangent of the circle they move on; they travel away from the centre *S* instead (!). This means the particles on the axis *SI* tend towards *I*, those on *SE* towards *E*, and those on *SA* towards *A*, "so that if there were not enough of them to occupy all the space between *S* and the circumference *AEI*, they would leave towards *S* all the space they do not occupy" (*adeò ut, si non sint satis multi ad occupandum omne spatium inter S & circumferentiam AEI, totum quod non occupant, relinquatur versùs S* — III, § 61).

Now Descartes says it is obvious that the space around *S* will be circular. His arguments are more than a little perplexing, but this need not concern us. What counts is that Descartes, in showing that the space surrounding *S* is "circular", has only "proven" that the *corpora solis & fixarum* are discs in a plane: his text explains by no means the spherical form of the sun and the fixed stars. Has Descartes tried to take advantage of the intuitive character of his diagram to make his reader accept a three-dimensional model? It is quite possible. At any rate, Pluche notices this shortcoming in Descartes' account and takes advantage of it, arguing that matter, if moved in circles or around an axis, would never arrange itself into a sphere (like the sun and the planets) but rather as a cylinder or a long spool (Pluche 1742, vol. II, p. 242[5]). Indeed: if we are to add, after Descartes' explanation, a third dimension, it could only be an accumulation of "slices" or planes as in fig. 2. This means that no matter how many slices we imagine (there can be as much as there are points on the axis around which the matter of the eddy revolves), the particles in those planes will always circulate around the centre of *their* planes and not around a *common* one as would be the case in a spherical model. Indeed, if the subtlest particles were to move towards the centre of the vortex they would find themselves lined up as points on an axis connecting the centres of multiple discs, which would in turn result in a cylindrical sun that would have the same length (in fact, height) as the axis. The shape of the earth would follow: heavier particles incrusting themselves on the borders of the eddy would develop into a cylindrical, elongated, "earth" in the form of a spool, and the other "planets" would be mere layers on the inside (or outside) of this cylinder (*Ibid.*, vol. II, p. 242-243).

Once again Pluche accomplishes a *reductio ad absurdum* of Descartes' TE: he adopts its starting conditions but by developing it more accurately than Descartes himself had done, he succeeds in showing that this TE does not lead to a world resembling ours and therefore does not deserve to be taken seriously.

If anything, Pluche's objections show that TEs, like real experiments,

---

[5] Pluche has added this example in the second (1740) edition of his *History*. Therefore it is absent from Freval's translation, which is based on the 1739 text.

are fallible[6]. By now however it should be clear that since Descartes himself intended the TE of the *World* to be an illustration (and legitimation) of his mechanist philosophy, Pluche's proof of its insufficiency had serious consequences for Cartesianism: his refusal to accept the identification of extension and matter (an essential condition for the success of the whole experiment) allowed him to call into question the entire concept of a "world-machine" (*machina mundi*).

3. Thought Experiments and the Rhetoric of Verisimilitude

When Georges Louis Leclerc, count of Buffon (1707-1788), starts working on the cosmogonic hypothesis that he will ultimately present in his *Proofs of the Theory of the Earth* (*Preuves de la théorie de la terre*, 1749), he abandons Cartesian mechanism in favour of Newtonian gravitation and its *actio in distans*. Nevertheless, he shares a Cartesian ambition: to establish, through the success of his cosmogonic TE, the legitimacy of his scientific method. But then again, true arguments, as we all know, are not necessarily persuasive — which means that Buffon cannot content himself with a TE that would merely be *true*. If he wants his experiment to convince his readers, he must convey it by means of a likely, i.e. *verisimilar*, story. This leads me to the following question: how does Buffon prepare his narrative so as to bear the burden of the rhetoric of verisimilitude?

3.1 Cosmogonic narratives as *exempla*

In the first volume of his *Natural History*, Buffon presents his cosmogony as an "essay", or example, of his scientific method. He writes: "we will give some examples of this method in the following discourses: the *Theory of the Earth*, the *Formation of the Planets*, [and] *The Generation*

[6] Sorensen calls errors similar to the ones Pluche points out, "oversuppositions": "The godlike power of stipulating tempts one to annihilate obstacles through a sheer act of will. However, if you assume too much, you may inadvertently trivialize the very problem the TE was intended to solve. Often, the trivialisation takes the form of circularity: the thought experimenter begs the question by presupposing what he aimed to prove." (1992, p. 257)

of the Animals". By doing so, Buffon echoes Descartes who at the end of his *Discourse on Method* announces three exemplary treatises, destined to show the usefulness of his philosophical method: the *Dioptrics,* the *Meteors,* and the *Geometry*.

The parallelisms between those two constellations of texts, however, involve more than the rather superficial observation that Buffon, after Descartes, has his methodological treatise followed by three essays. Indeed, we should also take into account the fact that Descartes' *Discourse on method* and Buffon's *First Discourse* both present a *new* method for "finding the truth in science". The "arts" of discovery (*artes inveniendi*) they present are destined to compete with, and ultimately to replace, already established scientific practices. As a result, these texts must be seen as instances of what rhetoricians call the "deliberative" *genus* or genre. In deliberative discourse, one interrogates oneself concerning the options that are to be taken in the future, and tries to establish, in accordance to the state of affairs currently accepted in judiciary (i.e., legislative) and epideictic (i.e. evaluative) discourse, a programme — in this case, a research programme (Hallyn 2000, p. 12).

This has some very interesting consequences for the status of the texts that accompany Descartes' and Buffon's treatises on method. For even if classical rhetoric considers the most persuasive *argumenta* of the deliberative genus to be, indeed, examples, rhetoricians have also pointed out that every example has a narrative structure, which should fulfil a number of criteria (if it wants to persuade, that is). And among the latter, truthfulness, as opposed to verisimilitude, is not an issue: "The public has to be lead to believe the story the author presents. Therefore, the ways of the story have to be plausible, or verisimilar" (Forestier 1996, p. 146).

It follows that any thought experimenter who wants to provide a persuasive — *mediative* — TE should relate the events he describes to a comprehensible system of causality, be it a logical or a doxical one. But since "doxical likelihood" involves a wide-ranging set of beliefs shared by the author and his public, persuasiveness and deductive soundness are two completely different things. As a result, the (perceived) plausibility and hence the discursive (argumentative) effectiveness of a mediative TE, is as dependent on the author's respect for the "encyclopaedia" of its public as on the quality of his deductions.

At this point, the author of a cosmogonic TE faces a particular challenge, because the substantiation of his claims may require him to call

upon scientific information his audience is not familiar with. Cases in point are the principles of Boerhaave's chemistry (as in Pluche's *History of the Heavens*) and the elaborate physical and mathematical explanations that underpin Buffon's speculations. But there is more. As a matter of fact, Pluche and Buffon knew very well that most of their readers were amateurs who received no scientific training whatsoever. How, then, did they manage to give their cosmogonic narratives their evidential character?

Basically, they both adopt the same strategy: through the extensive use of informative digressions they seek to expand, and sometimes even to reform, their readers' encyclopaedia. They try to add the rules and examples required to ensure the verisimilitude of their TEs, and to remove the rules that make them unacceptable.

An example will make this clear. It is well known that one of the most important objections Buffon had to deal with is the fact that he is unable to explain how the solar matter which, according to his hypothesis[7], formed the planets, did not fall back into the sun instead (which should have happened according to physics). Buffon tackles this objection in a quite remarkable way: he does not try to make his point by presenting some lengthy mathematical proof, but by presenting a series of analogies and comparisons destined to render his account more credible: continually accelerating rockets, eruptions of large volcanoes, "elastic forces", etc. In the end, he concludes:

> I admit that I do not know which one of the reasons I have just presented explains the change in the direction [and speed] of the planet's first movement, but those hypotheses are certainly sufficient to show that this change was possible, and even likely, and this is all I need. (1749, p. 143, my translation)

It goes without saying that this way of positing the "verisimilitude conditions" of TEs (i.e., by the thought experimenter himself) does not only have a constructive function. In the hands of a critic it becomes a

[7] His TE states, in short, that some 75000 years ago a comet crashed into the sun, projecting some of its matter into space. This incandescent gush of matter has been cooling ever since, and its condensed remnants have formed the planets of our solar system.

fearsome weapon: just remember how Pluche, in the *History of the Heavens*, attacked the Cartesian narrative.

To sum up, it may be said that the TEs presented by Pluche and Buffon are stories that have "abandoned the tacit conventions of the verisimilar, but have not yet reached the silence of [...] the *improbable*" (Genette 1969, p. 78, my translation). Their argument is "too remote from the commonplaces of the verisimilar to be able to rest on a broad consensus of general beliefs, and at the same time too dependent on the agreement of the general public to allow itself to expound, without further comment, actions which this public would, otherwise, not understand" (*Ibid.*, p. 79). Therefore, they "have to assure their own transparency, by multiplying explanations, by continually providing the general maxims the public is unaware of but that are required to understand them" (*Id.*).

3.2 Uncovering the Paradoxical Logic of Cosmogonic TEs

Genette (1969) and Forestier (1996) show that every narrative is built around a double structure of causality: a "chronological" chain where events determine their subsequent effects, and a "retrospective" chain where the state of affairs the narrator (the TEer) wants for the end of his story determines the causes that are brought into play. Moreover, both authors argue that the former chain of determinations is but a superficial effect of reading (or story telling), while the latter dictates the development of the story. Valéry sees it like this:

> What the reader sees as mechanical determinations has not been produced as such by the narrator. Of course, when he writes, *The marquise, desperate...* the author is not so free to continue with *...ordered a bottle of Champaign*, as with *took a gun and blew her brains out*; but in the real world things do not happen like this: when he writes, *The marquise*, he already knows whether the scene will end with a drinking-bout or a suicide, and it is in function of its end, that he chooses its middle. Therefore, and in contrast with what the reader's point of view suggests, it is not *desperate* that determines the gun, but the gun that determines *desperate*. (1950, vol. I, p. 1467, my translation)

This is equally true for the cosmogonist: if Buffon takes the crash of a

comet on the sun's surface as the starting point of his TE, it is because he believes that such an event is likely to lead to a world not unlike ours, which has all the characteristics he described in his *History and Theory of the Earth*. And if Pluche refuses to accept Descartes' "chaos" and "primitive matter", it is because he deems it impossible that such a state of affairs could lead to a κόσμος governed by divine Providence and created to fulfil each of man's needs — the world he himself described in the nine volumes of his *Spectacle de la Nature, Or Nature display'd* (1732-1750)[8].

The existence of this "paradoxical logic of fiction" led Genette to assign a "function" and a "motivation" to every narrative (we might add: argumentative) unit.

> If its function is (in general) that which a unit *does*, its motivation is what it *needs* to dissimulate its function. In other words, function is a profit, while motivation is a cost. Therefore, the efficiency of a narrative unit, or, if we prefer, its value, will be the difference offered by the subtraction: function minus motivation: V=F-M, we might call this Valincour's theorem. (*Idem*, pp. 97-98)

In order to illustrate how "Valincour's theorem" can be applied to a particular TE, I will now briefly consider Buffon's *Epochs of Nature* (1778).

Even if Buffon presents a cosmogony that is very respectful of observational data (the geological "archives" of our world), his ambition is to present a story with an obvious internal coherence and verisimilitude. However, in his attempt to combine factual truth, deductive soundness, and narrative cogency — an endeavour of which the success, especially with historical texts, is never to be taken for granted — he does not provide a strictly chronological presentation of the events he describes. Sure, his TE is divided in seven "epochs" but in almost all chapters flash-forwards (revelations on future events) and flash-backs (reminiscences of what has already been told), interrupt the chronological sequence. This causes quite some reiterations. Still, it would be an error to see this as a sign of gaucherie. For even if the mere repetition of an idea does not constitute an argument in its favour, it *can* bring about

[8] For a similar use of these concepts, see Hallyn (1999, p. 101 ff.).

persuasion. Therefore, all these reiterations and echoes can be seen as elements of a very carefully elaborated rhetoric apparatus: from the perspective of a "motivated" narrative which tries, as we have seen, to establish its own verisimilitude, flash-forwards and flash-backs allow the author to fulfil all the expectations he has created for his argument (which is the essence of the rhetoric of verisimilitude) and to confirm what he has already said.

3.3 Valincour's Theorem Revisited

As rhetorical parameters are quite important for the success of mediative TEs, Valincour's theorem deserves some refinement. For we could say, for instance, that the "value" of any section (sentence, paragraph) of a narrative (a TE, for instance) is determined by the uniqueness of its propositional content. Indeed, as the description of an event is repeated (announced, recalled), its various instances pertain less to the unfolding of the rest of the argument: its content can be found elsewhere. Nevertheless, we have seen that parallel passages can mutually motivate each other.

Therefore let x be the number of repetitions of the TEs "segment" of which we try to determine the argumentative value: the more it is repeated, the more it gains "auto-motivation" but loses functionality. Valincour's theorem becomes:

$$V = (F/x) - Mx$$

To quote Genette, "one should not laugh at this system of measurement, which is a but rude, but is well worth another one" (1969, p. 98). Indeed, the consequences of this reformation are, while obvious, very interesting. The more the propositional content of a segment is repeated, the more its narrative or argumentative value diminishes: it reaches zero and finally turns out to be negative. It becomes gratuitous, then redundant (from a scientific or philosophical point of view, that is).

But it is precisely to the extent that those sections become redundant that they can add aesthetic value to the TE. In the *Epochs of Nature* for instance, argumentation and narrative are occasionally put to rest. At those points, the TE's argument becomes a mere pretext to its author, who becomes an artist and undertakes the description of awesome scenes

of nature:

> Let us try to depict the extraordinary effects that have accompanied and followed the precipitation of those volatile substances, all separated, combined, sublimated, in the time of consolidation, and during the first cooling of the earth. The separation of the elements air and water, the beating of the winds and waters that fall all over a fuming earth, the depuration of an atmosphere that no ray of light has ever penetrated; that same atmosphere obscured again by dark smoke; a cohobation thousand times repeated, a continuous boiling of water, fallen down and immediately rejected, the washing of air, leaving behind unstable matters already sublimated, that separate themselves and fall down with more or less precipitation: What movements, what storms, have preceded, accompanied, and followed, the establishment of all those elements! (1778, p. 95-96, my translation)

PART II
THOUGHT EXPERIMENTS AND POSSIBLE WORLDS

Thus far I have only argued that Descartes', Pluche's and Buffon's cosmogonic TEs are *exempla* that purport to illustrate and legitimate their author's philosophical claims; for this reason I have focused on the fact that their argumentative efficiency resides in their perceived verisimilitude. Yet two highly important issues still need to be addressed: (a) these TEs are not *merely* verisimilar, because they claim to uncover the true constitution of the real world, and (b) even though TEs as a rule do not need to share all features of adequate referential discourse, those TEs attempt to. In the second part of this paper I will try to make these points by means of a possible-worlds analysis.

However, first I need to respond to an objection raised by Brown (1991) who is very sceptical about the usefulness of possible-worlds theories for the study of TEs. Indeed, he states that "talk of possible worlds is harmless" and that "it is even heuristically useful", but while he allows for quite some liberty in the positing of PWs this is also where he sees a major problem: they can not be inconsistent.

> [This] makes them inadequate for the analysis of a great deal of scientific reasoning. The simple fact is that many thought-experimental

> situations, like the scientific theories they deal with, are outright inconsistent. (1991, pp. 93-94)

If Brown's conclusion is to be seen, as I believe, in the light of the *ex falso aliquod* principle (according to which an inconsistent world would allow every possible inference and thus have no meaning whatsoever), the following constitutes a satisfactory response: (a) PWs merely *describe* states of affairs, and (b) paraconsistent and inconsistency-adaptive logics allow for a reasonable treatment of inconsistencies. This is why I could not agree more with Rescher and Brandom when they state that "inconsistent objects and worlds are feasible targets of rational consideration and scrutiny":

> They too can be meaningfully assumed, supposed, hypothesized, etc. And the supposition of such worlds is emphatically not an invitation to logical chaos. One can reason perfectly cogently and coherently about them. (1980, p. 4)

1. Possible Worlds as a Means for Analysing TEs

While every PW stipulates a state of affairs that allows the evaluation of a number of propositions, text-semiotic possible-worlds theories (which we need in order to tackle TEs) are characterised by the fact that they search for these propositions in texts. They assume that every text describes (and hence, constructs) a "world" and, conversely, that a "narrative world" only exists through its descriptions.

Texts that refer only to one state of affairs are rather uncommon: most of the time, they encompass various sub-"worlds": hypotheses, alternatives, dreams, desires... That is why the states of affairs described by texts ought to be represented not by isolated worlds, but by structured collections of worlds — "models". In other words: "the semantic domain of [a] text is not an individual world in a modal system, but [...] a system of worlds centred on 'the textual actual world'". Some of the worlds in this "system" function as "alternative PWs of the textual system of

reality" (Ryan 1991, p. 554)[9].

Although they are grounded on the same principles, there are important differences between the worlds and models of text semiotics and those of (modal) logic; I have studied them in my book *Les études littéraires et la sémiotique des mondes possibles* (p. 20 ff., *passim*). In short, we can say that as opposed to the Kripke- and Hintikka-type "first generation" theories used by modal logic, possible-worlds frameworks of the "second generation" are devised to study mental representations. However, for this purpose the constraints that govern the PWs of formal logic need to be loosened up: PWs of the second generation are not necessarily consistent (people's beliefs can be contradictory), they are not maximally complete nor closed under deduction (one might not be aware of the implications of his convictions) and their propositions are not necessarily numerable (it is very hard to enumerate someone's beliefs). Another very important feature of possible-worlds theories of the second generation is that the accessibility relation (which in modal logic is but a relation of co-possibility) is interpreted in terms of *conceivability*, and expresses (among other things) the relation between an author's views on the real world, the actual world of his text, and the textual alternatives to his text's actual world.

Ryan (1991a and 1991b) called upon this theoretical framework to elaborate two typologies of narrative discourse: the first distinguishes between various forms of "mimetic discourse"[10], the second takes into account characteristics of the accessibility relation in order to identify differences and similarities between textual PWs and representations of the real world.

For the first part of her analysis, Ryan (1991b, p. 24-25) stipulates

---

[9] From a cognitive point of view these textual PWs must be studied as dynamic networks of propositions, but it is equally useful to think of them as resulting from the application of a context-dependent valuation function to a context-dependent set of propositions. See my (2000), p. 76 ff.

[10] Ryan defines "mimetic discourse" as follows: "Under this label of *mimetic discourse* I understand utterance acts fulfilling the following conditions: (1) A mimetic utterance act makes singular existential claims ('there is an x,' rather than 'for all x'); (2) It describes particular facts and individuated entities; (3) it is proposed (really or in make-believe) as a version of a world existing independently of the discourse that describes it; (4) It is meant to be valued as either true or false in this world."

the following principles:
(1) The author of a text thinks about the world he lives in through a modal structure (a set of "worlds") that is seen as an adequate description of reality. This representation is centred on the actual world AW.
(2.) Every text describes a "universe" centred on an actual world TAW [textual actual world], which is presented as an adequate description of the "textual reference world" TRW. Moreover, TRW is presented (really or in make-believe) as existing independently of its description.
(3.) Every text has an implied speaker, i.e. the individual that "lives", so to speak, in the TRW and "fulfils the felicity conditions of the textual speech acts".

This allows for the following distinctions: (a) Is TAW similar to the actual world AW? (b) Does AW correspond to TRW? (c) Is the TRW adequately described by TAW? (d) Does the author (AS) "stand behind" the implied speaker (IS) to accept responsibility for his claims? In other words, do the real author and the implied speaker share the same beliefs? The results are presented in the following diagram:

	TAW=AW	AW=TRW	TAW=TRW	AS=IS
Nonfictional accurate discourse	+	+	+	+
Errors	-	+	-	+
Lies	-	+	-	-
Accidentally true lies	+	+	+	-
[Fiction	-	-	+	-]
True fiction	+	-	+	-
Unreliable narration in fiction	-	-	-	-

How do TEs fit into this system?
(1) As the TRW (the reference world) described by a TE is necessarily hypothetical (it is idealised, counterfactual...), the actual world of the model in which the TE takes place (TAW), cannot be the world we refer to when describing reality (AW); hence TAW $\neq$ AW — and for the same reasons, AW $\neq$ TRW.
(2) In the first part of this paper, I argued that TEs are legitimate philosophical devices only to the extent that they are carried out in enough detail to be repeated (to be falsified). This repetition, however,

can only take place if the TEer gave an adequate description of his argument — which makes TAW = TRW a necessary, albeit not sufficient, condition for the success of a TE.
(3) Obviously the TEer wants to reach a conclusion that will expand his knowledge of the real world, but as it is perfectly possible to do so while starting from idealised (Galileo on falling bodies), fictitious (Putnam's "twin-earth" experiments) or even counterfactual (Fogel's hypothesising about what could have happened if the USA's railroads had not been built) premises, TEs do not necessarily fail when their author can not guarantee the felicity-conditions of his narrative. Hence AS ≠ IS.

TAW ≠ AW, AW ≠ TRW, TAW = TRW, AS ≠ IS: this "profile" shows that TEs are indeed heuristic (illustrative, argumentative) *fictions*. Yet it is very interesting to note that some authors have gone to great lengths to ensure that their TEs are presented through a "true" story, i.e. a narrative organised around what they believed to be an adequate description of reality (in other words, the TRW of these TEs come *very* close to the real world). This is especially true for Buffon, who wants his *Epochs of Nature* to be a history of the world we live in even if he states that he is unable (and unwilling) to certify its truth[11]. One of the more interesting consequences of this situation is that his TE becomes an instance of the paradoxical category of "true fiction", which Ryan describes as follows:

> [A] fictional universe may be deliberately conceived and presented as an accurate image of reality. The difference between nonfiction and true fiction is that the former claims to represent reality itself (TRW = AW), while the latter represents a world TRW distinct from but very similar to AW. [...] True fiction exploits the informational gaps in our knowledge of reality by filling them in with unverified but credible facts for which the author takes no responsibility (as would be the case in historiography). (1991, p. 33-34)

As a matter of fact,

---

[11] The same thing goes for Pluche: while his criticism of Descartes' cosmogony is an example of "adequate non-fictional discourse", the Burnetian theory he advocates in the chapters on "Moses' physics" is but a hypothesis and therefore (he says) prone to the shortcomings that are typical of all human "conjectures".

> [t]he point of presenting the text as a fiction is that unverifiable facts can be directly asserted for TAW without being asserted for AW, and therefore without compromising the credibility of the author. (*Ibid.*, p. 35)

Ryan's second classification of fictional texts follows from a thorough analysis of the characteristics of the accessibility relation[12] (in the above-mentioned sense of conceivability relation) that holds between the author's AW and his text's TAW. In this typology, the central issue is "what properties of his AW does the author keep, or alter, when he stipulates the PWs of his text?" Here are the properties she takes into account:

A. *Identity of properties*: the objects common to TAW and AW have the same properties.
B. *Identity of inventory*: the same objects furnish TAW and AW.
C. *Compatibility of inventory*: TAW's inventory includes all the members of AW, as well as some "native members".
D. *Chronological compatibility*: TAW's present is not posterior in absolute time to AW's present, which means that we need no temporal relocation (towards the future, that is) to contemplate the entire history of TAW.
E. *Physical compatibility*: AW and TAW share their natural laws.
F. *Taxonomic* compatibility: AW and TAW contain the same species, which are characterised by the same properties.
G. *Logical compatibility*: TAW respects the principles of non-contradiction and of excluded middle (as does AW).
H. *Analytical compatibility*: TAW and AW share analytical truths, i.e. if an object of AW is referred to in TAW, it keeps its essential properties.
I. *Linguistic compatibility*: the language in which TAW is described, can be understood in AW.

These definitions can be found in Ryan 1991b (p. 32-33); once again,

---

[12] Ryan speaks of "relevant types of accessibility relations". As it is possible however to assign various characteristics to *one* accessibility / conceivability-relation, there is no need to assume several conceivability relations.

they allow the construction of a comparative tableau:

	A	B	C	D	E	F	G	H	I
Accurate nonfiction	+	+	+	+	+	+	+	+	+
True fiction	+	+	+	+	+	+	+	+	+
Realistic & historical fiction	+	-	+	+	+	+	+	+	+
Historical fabulation	-	-	+	+	+	+	+	+	+
Realistic fiction in "no-man's land"	-	-	-	+	+	+	+	+	+
Anticipation	+	-	+	-	+	+	+	+	+
Science fiction	+/*	-	+/-	-	F+ F'-	+/-	+	+	+
Fairy tale	*	-	-	+	-	-	+	+	+
Legend	-	-	+	+	-	-	+	+	+
Fantastic realism	+/*	-	+/-	+	-	+	+	+	+
Nonsense rhymes	*/-	-	-/+	#	-	-/+	-	+/-	+
Jabberwockism	*	-	-	#	-	-	?	+	-$
Sound poetry	*	-	-	#	-	-	-	-	-

"*" means "non-applicable because of a '-' on C"; "#" is "non-applicable because of a '-' or '?' on G", and "-$" stands for "incompatibility restricted to most nouns and verbs". (Ryan 1991b, p. 34)

2. What Function for Thought Experiments?

When I wrote that Descartes, Buffon and Pluche "went to great lengths to ensure that their TEs are presented by means of a 'true' story" I meant that they wanted to present narratives which, according to their contemporaries' standards, respected the criteria *A* through *D* as well as *F* and *G*. Regarding *E* (physical compatibility), *H* (analytical compatibility) and *I* (linguistic compatibility) however, they adopted another stance. For they did not devise their TEs in an already accepted paradigm; instead, they believed that the success (i.e., the cogency) of their TEs explaining the formation of our world would

turn out to be justifications of the new theoretical (conceptual) apparatuses they developed. What they felt was at stake here, was the appropriateness of their hypotheses towards providing a means to study our world (*E*), the legitimacy of the secular logic that governs modern science (*G*), and the soundness of the conceptual apparatuses they used to present their TEs (*H-I*). As a result — and this is, of course, quite paradoxical — the success of their respective TEs "proved" the soundness of Cartesian physics as well as Newtonian mechanics, the atheistic logic of Buffon as well as the providentialist account of Pluche... but this brings us back to our starting point: the eminently rhetorical function of these TEs.

It should be clear by now that the deliberative (illustrative and argumentative) function TEs have in Descartes' and Buffon's work, is quite different from the critical function Kuhn assigned to them:

> Historically their role is very close to the double one played by actual laboratory experiments and observations. First, TEs can disclose nature's failure to conform to a previously held set of expectations. In addition, they can suggest particular ways in which both expectation and theory must henceforth be revised. (1977, p. 261)

According to Kuhn, TEs show the insufficiencies of an already established paradigm (i.e., of an already established set of concepts and theories). By doing so, they play a crucial role in the "reconceptualisation" process that characterises scientific revolutions. This is clearly not the function of TEs in Descartes' and Buffon's work. On the contrary: their arguments attempt to show the adequacy of *new* paradigms which are to be preferred for strictly philosophical reasons (e.g., "secular science is more valuable than theology-inspired 'physics'"), not because they give better accounts of reality. At least in this sense, the TE Pluche thinks of when he criticises Descartes' cosmogony has a more classic, or "kuhnian", purpose: in Pluche's opinion, the failure of the TEs that have been put forward by various cosmogonists proves that mankind will probably never have access to the insights necessary to fully understand the origins of our world. The laws of physics, he says, can only explain the world's continuing existence (as Newton's had already stated); if we want information

about its origins we should turn to the Bible.

Universiteit Gent

REFERENCES

Blumenberg, Hans (1999), *Die Legitimität der Neuzeit*, Frankfurt am Mein: Suhrkamp.

Brown, James R. (1991), 'Thought Experiments. A Platonic Account', in Tamara Horowitz & Gerald J. Massey (eds.), *Thought Experiments in Science and Philosophy*, Savage: Rowman and Littlefield, pp. 119-128.

Brown, James R. (1991), *The Laboratory of the Mind. Thought Experiments in the Natural Sciences*, London & New York: Routledge.

Buffon, Georges-Louis Leclerc (1749-1798), *Histoire naturelle, générale et particulière, avec la description du cabinet du roy*, Paris: Imprimerie royale, 31 vol.

Buffon, Georges-Louis Leclerc (1988), *Époques de la Nature*, Paris: Éditions du Muséum (Jacques Roger ed.).

De Baere, Benoît (2000), *Les études littéraires et la sémiotique des « mondes possibles ». Une réflexion critique sur l'état actuel des recherches*, Ghent: Communication and Cognition.

De Baere, Benoît (2001), *Trois introductions à l'Abbé Pluche. Sa vie, son monde, ses livres*, Genève: Droz.

De Baere, Benoît (2004), *La pensée cosmogonique de Buffon. Percer la nuit des temps*. Paris: Honoré Champion.

Descartes, René (1996), *Œuvres complètes*, Paris: Vrin, 11 vol. (Adam & Tannery eds.)

Dole_el, Lubomir (1998), *Heterocosmica. Fiction and Possible Worlds*, Baltimore & London: Johns Hopkins.

Fogel, Robert W. (1964), *Railroads and American Economic Growth: Essays in Econometric History*, Baltimore: Johns Hopkins.

Forestier, Georges (1996), *Essai de génétique théâtrale. Corneille à l'œuvre*. Paris: Klincksieck.

Genette, Gérard (1969), 'Vraisemblance et motivation', in *Figures II*, Paris: Seuil, pp. 71-99.

Hallyn, Fernand (1999), 'Descartes et la méthode de la fiction', in Jacques Berchtold & Christopher Lucken (eds.), *L'orgueil de la littérature. Autour de Roger Dragonetti*, Genève: Droz, pp. 89-106.

Hallyn, Fernand (2001), 'De l'anagramme au cryptogramme', in *Regards sur*

Galilée. Actes de la journée Galilée, Lille: Institut de Recherches sur l'enseignement des mathématiques (Université Lille II), pp. 15-29.

Hallyn, Fernand (1999), 'L'illusion et l'erreur dans le Monde de Descartes', in *Cuadernos de filología francesa* **11**, Cáceres: Facultad de filosofía y letras (Universidad de Extremadura), pp. 37-50.

Hallyn, Fernand (2000), 'Une rhétorique sans frontières', in *Théorie, littérature, enseignement* **18**, pp. 7-27.

Hofstadter, Douglas R. & Dennett, Daniel C. (1982), *The Mind's I*, New York: Bantam.

Irvine, Andrew D. (1991), 'Thought Experiments in Scientific Reasoning', in Tamara Horowitz & Gerald J. Massey (eds.), *Thought Experiments in Science and Philosophy*, Savage: Rowman and Littlefield, pp. 149-165.

Kuhn, Thomas S. (1977), 'A Function for Thought Experiments', in *The Essential Tension. Selected Studies in Scientific Tradition and Change*, Chicago & London: Chicago UP.

Norton, John (1991), 'Thought Experiments in Einstein's Work', in Tamara Horowitz & Gerald J. Massey (eds.), *Thought Experiments in Science and Philosophy*, Savage: Rowman and Littlefield, pp. 129-148.

Norton John D. (1996), 'Are Thought Experiments Just What You Always Thought?', *Canadian Journal of Philosophy* **26** (n° 3), pp. 333-366.

Plett, Heinrich F. (1979), *Textwissenschaft und Textanalyse. Semiotik, Linguistik, Rhetorik*, Heidelberg: Quelle & Meyer.

Pluche, Noël-Antoine (1742), *Histoire du ciel où l'on recherche l'origine de l'idolatrie et les méprises de la philosophie, sur la formation des corps célestes & de toute la matière*, Paris: Veuve Estienne, 2 vol. (3rd augm. ed.).

Pluche, Noël-Antoine (1741), *The History of the Heavens, Considered according to the Notions of the Poets and Philosophers, and Compared with the Doctrines of Moses*, London: Osborn, 2 vol. (transl. from the 1739 edition by J. B. de Freval)

Rescher, Nicholas & Brandom, Robert (1980), *The Logic of Inconsistency: A Study in Non-Standard Possible-Worlds Semantics and Ontology*, Oxford: Blackwell, 1980.

Ryan, Marie-Laure (1991a), 'Possible Worlds and Accessibility Relations: A Semantic Typology of Fiction', in *Poetics Today* **12** (3), pp. 553-576.

Ryan, Marie-Laure (1991b), *Possible Worlds, Artificial Intelligence, and Narrative Theory*, Bloomington: Indiana University Press.

Shea, William (1991), *The Magic of Numbers and Motion. The Scientific Career of René Descartes*, Canton: Watson (Science history publications).

Sorensen, Roy A. (1992), *Thought Experiments*, Oxford & New York,

Oxford UP.
Toulmin, Stephen (1990), *Cosmopolis. The Hidden Agenda of Modernity*, Chicago & London: Chicago UP.
Valéry, Paul (1950), 'Fragments des mémoires d'un poète', in *Variété*, *Œuvres*, Paris: Gallimard.