A GLIMPSE

BEYOND

SEARCH

ENGI

NE

S

CHRISTOPHE BRUNO

回

Verb thou art,
and unto verb shalt
thou return

I N the present paper, I describe briefly some projects (that do not exist as yet) which one can see either as possible utilitarian global structures or as art-fiction pieces subverting these structures. I'm not going to distinguish between the two, as their development is not advanced enough. These projects are in progress on iterature.com and you will just find a foretaste here.

A brief history of global narcissism

The Web has often been compared to our own memory, or to con-sciousness. There are indeed some similarities, simply because they share the same raw material: language. As any language-related phenomenon, the Web is on the one hand a natural thing for human beings, and on the other hand it partakes of the most absolute otherness. Like any language-related phenomenon, it can sometimes liberate us, but most of the time, it seems that we are locked in this prison made of words, more tangible than any mate-rial prison. Paradoxically, it looks wide open to the world, through the media. This makes the question of escape more problematic; and indeed why escape? For some of us, this situation is quite com-fortable, for others, it's just unbearable. It is even more complicat-ed because the media send us back an image that we feel intimately connected to ourselves, like a global mirror.

More than two hundred years after Jeremy Bentham described his Panopticon[1], a prison system where everybody can watch over everybody (but some watch more than others), the Web has

become the emblematic implementation of this media prison with no definite border, no separation between the inside and the outside.[2] Bentham's utilitarian philosophy discussed the question of tracking information, reducing reality to elementary parts and then trying to describe the general dynamics of causes and effects, for the achievement of what may be called a global causal machine, or, after the title of one of the Pentagon's projects, the *Total Information Awareness* engine.[3]

Recent media events concentrate the culmination of this issue, such as reality-TV and 9/11. It looks as if the world had gone through the looking-glass, enduring some kind of phase transition. In fact, the concept of mirror appears to me as related to the question of globalization of thought: both of them show an aspiration towards some phoney ideal of unification through the use of media, as suggested in the conceptual Google Art piece *Self-portrait* by Valéry Grancher.[4] This is actually all I'm going to talk about: how Web global structures, like Google, deal with symbolic forms.

Thus, before focusing on the Web, I believe it may be instructive to insist a little bit on this question of mirrors and globalization by sketching a short and careless picture, à la Google, of what could be the beginning of a history of global narcissism (some of these historical landmarks will accompany us all along this article):

… 1787… Jeremy Bentham imagines the Panopticon or Inspection house…

… James Joyce invents the concept of "epiphanies", one of the first literary form involving the issue of globalization… walking down the streets of Dublin and collecting bribes of speeches from different people and different contexts…

… George Orwell, and later Philip K. Dick, describe the consequences of a totalizing knowledge…

… science looks at itself… Bertrand Russel questions the very roots of mathematics with his famous paradox, followed by Kurt Gödel opening the field to Alan Turing and the founding of computer science[5]…

… concentration camps… a part of mankind is designated as the ulti-

mate cause, then they are numbered and treated as refuse… This logical perversion - or accomplishment? - of the Panopticon, builds the global nature of mankind on the roots of its radical division, putting away what we don't want to see, our very image, the dark side of the Ideal…

… Jacques Lacan promotes the *mirror stage*…

… Arthur Danto… art looks at itself…

… in 1969, the Earth looks at itself on TV, filmed from its alter ego, the Moon…

… 9/11, because of the choice of the target - the Twin Towers, establishes a link between global terrorism and the concept of mirrors, duplication and memetics…

…

Recently, a frightening murder took place in France: an old woman was killed and was discovered with a microphone inserted into her vagina. This crime provides the most striking definition of what we call media. It can be seen as the missing link between concepts such as *Total Information Awareness* (T.I.A.) and questions such as femininity and castration, which now resurface in their most pornographic religiosity and inquisitorial form in the global snuff movie we are living. Curiously enough the old woman, who was well-known and loved in her city for her kindness, was nicknamed "supergranny".

We shall now leave behind us these considerations and turn to some superpowers of the media world.

Through the searching-glass: Google & al.

The attempt of extracting our most intimate thoughts and desires by all means has become a well-known strategy of large scale marketing. This strategy is obvious in the last product of Google, Gmail, which is problematic because it is a threat to email privacy, even though only bots have theoretically access to the content of the mails. The year before, Google had bought Blogger and many people were puzzled by this, because Google still had the image of the "nice guys" coming from the Linux world and the libertarian

side of the Web. But, as everybody knows now, the reason why they bought Blogger is that this is an extraordinary tool to track the behaviour of internet users and to optimize their advertising system of Adwords-Adsense.

There is an important distinction to make between Google and other companies like Microsoft, Yahoo or other portals. From the beginning, Google's ethics was to refuse to adopt the traditional business model which mixes advertising with the search results because it leads to clashes of interests and cheating the user.[6] Google is not a *spectacle provider* that imposes a brand, a particular content or an ideology. It draws its strength from its withdrawn position, from the spectacle provided by the set of all the content publishers: mere internet users, bloggers, messageboards, newsmagazines, etc… i.e. potentially mankind as a whole. The Adwords system taxes the circulation of browsing, the flux of desire, and it has to respect the speech of the internaut as much as possible because this speech is its raw material. In the same way as global companies now incorporate environmental arguments in their marketing discourse, Google has become the promoter of an ecology of speech. But the other face of ecology is vampirisation. Like a vampire, Google needs the blood of living people and like a vampire, it has no image—hence its minimal interface. This position was essential for a performance like the *Google Adwords Happening*[7] to occur; it wouldn't have been possible on MSN or Yahoo because they did not go far enough in this ultimate stage of capitalism, as Jeremy Rifkin would say.

Other companies are entering the game. IBM's WebFountain[8] consists in an army of hardware and software, dedicated to one purpose: making sense of the ocean of information, opinion, and falsehood that continually percolate on the Internet. It is a Webscale mining platform that extracts trends from massive amounts of text. What is the limit to this tendency towards the panopticon? Just have a look for instance at projects like *reality mining*,[9] they will give you a foretaste of our future. Personally, I don't see any clear limit. Capitalism needs more and more information to opti-

THE WORLD ACCORDING TO SOFTWARE

mize its processes and consumers are fond of such services…
as long as they respect their private life of course. But the point is
that the border of our intimacy is moving very fast, and not only
because of war on terror: for instance one of the effects of the
whole reality-TV industry is to promote exhibitionism and gradu-
ally bring people to share their private feelings, so that at the end,
the private sphere is dramatically shrinking, and people are sup-
posedly happy.

In what follows, I will adopt the following strategy that consists in
identifying with the "nice guys" and try to extrapolate what exists
today, as far as global structures are concerned. Indeed, the real
core of the web is these global structures: search engines, virtual
communities, Friendster, Adwords-Adsense, news aggregators,
Web-mining tools, weblogs, etc. All these textual or relational
structures aim at implementing new Internet symbolic protocols,
which complete or transgress the usual technical Internet proto-
cols (I mean the technological core of the Internet that can of
course also be seen as a symbolic form).

I believe one of the reasons why Google is the most representative of
all these global structures is because they were the first to deeply
understand the phase transition that occurred, transforming our
world into a global mirror, behind which they took position to
watch on us. And another reason is that they chose the kingdom of
words as their kingdom.

The Global Text

Capitalism has reached some new stage of development in which
words, our raw material as speaking beings, are now revealed as a
commodity ("Generalized Semantic Capitalism" as I called it). In
the Adwords system, each word is associated to its price, which is
related to its popularity. But the fact that words have a price is also
astonishing just because one relates a word with a number. In the
same way as markets are objects of scientific study, language itself
is becoming the object of a new science, a sort of interbreeding
between linguistics and the mathematical study of dynamic sys-

tems. Of course, as we were born and die embedded in language, any attempt to approach language as an external object can be considered as misleading. But so is linguistics.

- One of the main characteristics of the signifier is its differential aspect, introduced by Ferdinand de Saussure. Therefore, what we should be interested in is not only the value associated to one word, but the correlation between two (or more) words, i.e. the probability transition to go from one word to others. The starting point for studying such matter is search engines. In relating words to words and texts to texts, search engines unify the Web to form what I call the Global Text.[10]

- Search engines like Google don't necessarily care about who you are as an individual, but more about what you think or how you act in given circumstances: if somebody reads a sentence on the web, what advertisement should they associate to this sentence so that this person will be most likely to click on the ad etc… if somebody has thought A and now thinks B, what is the probability to think C?

- This situation has an equivalent in the field of exact sciences, which is called *change of variable*. For instance, if you try to understand the interaction between two clouds, you will soon realize that you need to understand more about the structure of a cloud in term of its molecules. Thus you switch from the variable *cloud* to the variable *molecules*.

- In terms of identity, we have switched from variables such as *individuals* to variables such as *elements of discourses*. The individual is then seen as a meta-object, integrated over his/her path in the space of discourses. For instance, the entity whose speech is the sum of all the speeches of mankind was called Gogol in the piece *GogolChat*[11] (this somehow looks like a kind of *ergodic hypothesis*: imagine that the whole set of texts of the Web had been written by only one person but at different moments of his/her life, instead of many persons at some arbitrary moments).

- This Global Text is our object of study. The question is how to extract hidden information from this huge textual soup, to grab language and its structure from the outside (as I said, it's misleading) and to

THE WORLD ACCORDING TO SOFTWARE

understand what it looks like *at large scale*, as if we had gone very
far away from it, turned back and seen it as if it was a landscape.

Topological mining: what is the large scale structure of language?

I'm not going to answer the question, but just try to explain why
I believe it has some relevance.

While a hyperlink relates a word to one page, search engines relate
a word to many pages: the list of pages you get when you perform
the search on this word. These search engines are the real browsers
to navigate on the web. Imagine the following experiment: on
a Web page, select a word, use it as input in a search engine, and
get the list of texts which contain at least once the word you typed.
Now, as a second step, choose one of these texts randomly, and in
this text, choose a word, also at random, different from the first
one. Then reiterate the first step. In fact, any word (or sentence)
can lead to any other word (or sentence) and in multiple ways, but
with different probabilities. In this way, you would navigate trans-
versally across the Global Text, in a manner which is closer to *asso-
ciative ideas* than to the usual utilitarian use of search engines.
These processes are quite close to those employed in art pieces
such as *TheBot*[12] or *GogolChat*.

Search engines, just as language, have a utilitarian side and a non-utili-
tarian one: when you speak, you are subjected to many constraints
(related to grammar, meaning, time, identity etc…) which are
imposed by the fact that for your interlocutor to understand you,
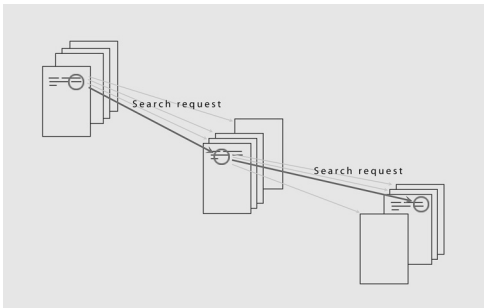
there has to be some normative protocol; this is the utilitarian aspect. The non-utilitarian side is glimpsed in dreams, associative ideas or poetry but is the hidden part of the iceberg, operating underground and despite the will of the subject.

The non-utilitarian kind of browsing I described reveals a very general structure, that mathematicians call hypergraph (a hypergraph is a generalization of a graph whose edges connect more than two vertices, instead of each edge connecting just two vertices).[13] On this hypergraph, two words (or group of words) will be more correlated when they appear both in the same texts more often (you can have an approximate idea of the magnitudes of these correlations if you look at the number of search results for these two words divided by the product of the numbers of search results for each one of the words). Then you realize that there are some preferred paths which are statistically more probable; each thought association has a probability of occurrence; you have high-flux zones and also regions of low correlations.

In the framework if our approach, if it was possible to draw this hypergraph we would get a global mapping of language, showing its topological configuration. We could in principle also derive quantities such as curvatures and distances. This is a very difficult task since it involves highly non-trivial mathematics and non-local calculations, involving the full information contained in search engines and not only some of the correlation functions
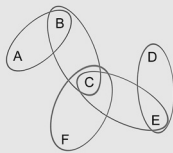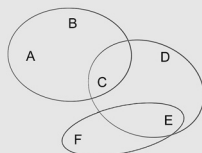
Example of graph   Example of hypergraph

THE WORLD ACCORDING TO SOFTWARE

(there are actually a lot of studies about the global geometry of the Web seen as a graph and locally defined by its hyperlinks structure, but I found none involving search processes as I defined above.[14])

- Furthermore, it seems very difficult to give any interpretation of this hidden structure. The slip from one expression to another one in its neighbourhood may be related to metonymy. The situation where two very de-correlated words are both strongly correlated to a third one seems related to some metaphorical processes. Unfortunately, the interpretation of the main information, the topological structure of the whole hypergraph, still remains unclear.

Dreamlogs or discourse-disentangling engine

After *Topological mining*, let's consider a more concrete approach. Imagine that you want to follow these thought associations that run across the Global Text in a transverse way, browsing from text to text using discursive neighbourhoods and not sticking to the linearity of one text. You are led to the concept of *Dreamlog*. For instance, you start with an element of discourse, and little by little you perform search requests and let the propagation occur, as defined above, until you reach some other discursive position, the negation of the point of departure for instance. Within the *ergodic hypothesis*, it's a little bit like when you have some convictions as a young person, and then, growing up, you slowly realize that you were wrong and that it is just the opposite.

Because two opposite discourses are put in relation, this concept attempts to reveal associative processes which may be non-local: they involve language at large scale. In that sense, *Dreamlogs* are the perfect tools for the study of discourse at the age of globalization. Interestingly enough, although we started from the so-called non-utilitarian side of language, *Dreamlogs* actually have a utilitarian aspect. Theoretically, if you can track all thoughts in such a way, you should be able to make some statistical predictions about

thought processes and therefore about our behaviour as consumers. We shouldn't be surprised at this: there is no definite border between non-utilitarianism and utilitarianism. As privacy vanishes, the empire of totalitarian knowledge expands and may transform anything into a commodity or a means of production. Therefore the Web shows a commodification of thought and what we once believed to be in the area of dream, poetry and art, will sooner or later cross this thin membrane that separates the two opposite fields.

Live speech

The concepts I describe above deal with the Text as a dead medium, even though it is fluctuating and ever changing. Although we can imagine that we could follow the evolution of language in real-time, they look like more an autopsy or archaeology of language.

This is not the case in the Google Adwords system where the dimension of speech literally perforates the web. What is quite unique in this Adwords service is that speech is a priori uncensored (censorship arises in a second step and is regulated by a computer program) and can be targeted to hit any user request, any user thought, addressing potentially all mankind. Many net.art performances involve the question of speech in quite a radical way: *Toywar*[15] by etoy.com was subverting search engine requests by playing with the equivocation etoy/etoys. The *Google Adwords Happening* generalizes this approach to all language: each (key)word is intercepted by an enigmatic ad which interrupts the resolution of meaning. In that sense, these performances put on stage the introduction of the subject of speech into some textual medium, paving the way to incompleteness and radical surprise, which of course strongly collide with the iron rules that sustain the utilitarianism-oriented social networks.

In what follows, I will distinguish between two distinct dimensions in speech: first the continuous reticular dimension, that uses the hypergraph's a-temporal structure mentioned above. The second is the dimension of the interruption, or scansion, which is induced by

the presence of the Other and is related to concepts such as inhibition or censorship.

Both dimensions reveal a lack of knowledge and control on our own speech but in different ways. This lack of knowledge has its own rules: our speech is not arbitrary even if we don't control it. To reinforce the dimension of live speech on the Web, we can imagine different mechanisms. I will here describe two of them: the *Slip-of-the-tongue engine*, and the latency engine.

Slip-of–the-tongue engine

Automatic production of text is expanding on the Web. Not only in the old way of the text generators, but more as semantic viruses that tend to populate the Web. Spambots generate texts to cheat on spam filters. Other programs generate fake WebPages to deceive search engine spiders and to improve website rankings. Nowadays, more than two third of all e-mail messages are spam and a large part of it is automatically generated. One may easily imagine that some day, the main bulk of mail and WebPages will be constituted by computer generated text[16].

If we were able to build the analysis engines described above, *topological mining engine* or *Dreamlog*, we could use them as generative tools to produce signifier according to the topological laws of the Global Text.

The *Slip-of-the-tongue engine* would be a hybrid of *Dreamlog* and chat in which the writings of each user are expanded or transformed according to these laws. The speaker loses control of the emission of speech, giving rise to some more reticular modes of communication (the experiment is even more interesting when the speech is transformed while the speaker is not aware of that, this possibility being actually allowed by the protocols of the Web).

Losing control of one's own speech is probably one of the most intriguing experiments one can make as a speaking being, since it underlines a tension between our unity as an individual and our dislocated being in terms of patchwork of elements of discourse.

Latency engine

Puritanism keeps revealing its spectacular dimension: in the beginning of 2004, Janet Jackson's breast became the most seen media icon ever, surpassing 9/11, as if media bred its own perversion, duplicating what is supposed to be hidden in a quasi-algorithmic recursive phenomenon. The desire of a nation suddenly erupts, hiding behind indignation: this cannot be the image of the USA. About one month before, the same spectator was inspecting the inside of Saddam Hussein's mouth. On the video one could see the latex-gloved hands of the medical officer holding the flash light. Faced with the latent threat of the T.I.A., "I will track your guilt down the interior of your body", the "nipplegate" brings its own answer, scandalous and totally unexpected.

What happened exactly between the two events? It would be the job of the *Latency engine* to discover hidden relations, to analyse what was repressed in the first event and suddenly resurfaced after some time. It is more powerful than a *Dreamlog* because here, both events appear at first as enigmatic, cut from their symbolic context. The *Latency engine* would cultivate signifiers around the traumatic event, reconstructing the material left pierced. The signifiers would be fostered by some automatic processes, generating massive amounts of WebPages with some calculated content and submitting them to search engines. These non-local viral propagation processes would somehow mimic the mechanisms that take place in our unconscious. We usually have the idea that *memes* replicate themselves identically. However, memetics is not the only mode of propagation for thought. In our minds, censorship constantly watches over our thought processes, but this doesn't mean that censored processes stop being active. On the contrary, their unconscious work plays an important role in our life, and they often resurface, once transformed, in some unconsciously driven phenomenon such as dreams, slips of the tongue etc… The *Latency engine* would provide such mechanism on the web, which would allow some components of the censored speech to continue an underground life in the expectation of a future resurgence.

Although we are very far away from their implementation, it's probably worth having in mind this sort of machines which could be used as simulating tools for emerging disciplines such as trend analysis.

Logical time, Turing Machines

So much for the reticular part of speech. We shall now end this paper with an evocation of the dimension of scansion.

Imagine now that a machine is built, which would indeed record not only any non-human related physical event but also any thought, any speech, any human action, a benthamian-like machine, with which lies would be impossible: the T.I.A. engine as we said before. We can then address the following classical question: how predictive can such totalizing knowledge be? For instance, as Philip K. Dick imagined that people might succeed in predicting crime just before it happened, or just as Google wants to predict the number of internet users who are going to click on an ad.

This problem has already been addressed in many forms in the history of thought. In 1845 Edgard Allan Poe in *The purloined letter*[17] gave an indirect answer to Bentham. The short story puts on stage the failure of totalizing knowledge in the resolution of a police enigma. An important letter has been stolen by a Minister. The police know that he hides the letter in his house and they are in charge of getting it back. The articulation of the story is as follows:

·   The police *think* that the Minister will hide the letter so that the best scientific police will have the least chance to find it.
·   The Minister *guesses* that the police will *think* this way and just leaves the letter slightly made-up and torn on his table. Indeed, the police search the whole house, inch by inch, but can't find it.
·   The detective Dupin *understands* that the minister has *guessed* what the police *think* and finds the letter as soon as he enters the house.

The resolution of the enigma by Dupin uses identification strategies, as in the odd/even game, instead of systematic search. This identi-

fication process assumes that the player has to guess the thought processes of the adverse part, and this process in turn potentially involves the same guessing strategy, in a mirroring play of anticipation. Each adversary has to advance his pawn while lacking knowledge about the adverse strategy. Each decision becomes part of the input, in a recursive way. As a result, the lack of knowledge, and not its completion, becomes the motor of the decisional process, implying a structural lack of predictability (which doesn't contradict the existence of a certain effectiveness of predictability, both in science and in occultism!).

This short story was commented by Jacques Lacan in 1956.[18] Its problematics is actually strongly related to another of Lacan's texts, "Logical time and the assertion of anticipated certainty"[19], from 1945, in which he discusses a logical enigma called the *prisoner's dilemma*.[20] It's quite noticeable that this family of logical games was also investigated in parallel in the late forties by the founders of computer science.

In case you don't know it, the famous article by Alan Turing, *Computing Machinery and Intelligence*,[21] describes a variant of the *prisoner's dilemma*, called the *imitation game*, between a man and a computer, also known as the *Turing test*. "… the 'imitation game' […] is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. The interrogator is allowed to put questions to A and B…". The question is then "What will happen when a machine takes the part of A in this game?"

There is a huge amount of writings about these topics, and it has become a staple of computer science, game theory, economy, or sociology. You will find a lot of references on the Internet as well as some online simulations.[22] It would be much too long to discuss these complex matters in detail here. My aim in this introductory

paper was just to pinpoint the relation they have to the answers by Poe and Lacan to Bentham's panopticon, about how to prevent the mithridatization of human speech by pervasive science. I leave the description of a possible *Logical-time engine* for further studies. To end this paper, let me just give an idea on how globalization has changed the parameters of the game. Firstly, although the T.I.A machine doesn't exist, the threat of its implementation is much more present than a few years ago. This of course has an impact on the way subjects deal with the anticipation of the effect of their acts: if you know you are being spied on, you don't act in the same way. Secondly, it suddenly seems that acts have a value only if they are broadcast worldwide, which is problematic when a subject is supposed to wait for the feedback of such an action. Feelings of powerlessness massively propagate, while (h)ac(k)tivists focus on global performances playing with the questions of fakes and equivocation (which is precisely what the T.I.A. machine wants to eradicate), and while global powers elaborate anticipative worldwide strategies and launch pre-emptive wars. Thirdly, as panopticon expands with the blogosphere and reality-TV, the concept of act itself becomes more and more radical: from Abu Ghraib prison's blogging torture to suicide bombing, the ultimate act, on which the T.I.A. machine has no control

.

1. *http://www.ucl.ac.uk/Bentham-Project/info/wwwtexts.htm*

2. About Panopticon vs Internet, see for instance: Glicenstein, J.: "Le paysage panoptique d'Internet, remarques à partir de Jeremy Bentham", *Autres sites, nouveaux paysages*, Revue d'esthétique n° 39 (2001), p. 101.

3. The T.I.A. is a project by the Pentagon. Originally *Total Information Awareness*, then rebaptised *Terrorism Information Awareness*. More information on: *http://en.wikipedia.org/wiki/Total_Information_Awareness*

4. *http://www.nomemory.org/data/selfportrait.html* by Valéry Grancher, 2002

5. The proof of Gödel's first theorem consists in constructing the statement "*p* = 'This statement cannot be proven'" within a formal axiomatic system. As such, it is related to the liar paradox. Although it's always dangerous to mix high level mathematics with common knowledge, let's say that it has left us with the question whether human thought may or may not be comparable with a Turing machine because of self-referencing loops that would arise from set theory paradox (I don't claim this last sentence is actually meaningful, but this is the common idea that people have of Gödel's theorem implications).

6. *http://www-db.stanford.edu/~backrub/google.html* Appendix A

7. *http://www.iterature.com/adwords* by Christophe Bruno, 2002

8. *http://www.almaden.ibm.com/webfountain*

9. *http://reality.media.mit.edu/sna.html*

10. In relation to this, see also the *Universal Page*, by Alexei Shulgin and Natalie Bookchin in 2000: *http://www.walkerart.org/gallery9/universalpage*

11   *http://www.iterature.com/gogolchat*
    by Christophe Bruno & Jimpunk, 2002

12   *http://thebot.plagiarist.org* by Amy Alexander, 1999

13   *http://en.wikipedia.org/wiki/Hypergraph*

14   I'm still searching though

15   *http://toywar.etoy.com* by etoy.com, 1995

16   Examples of automatically generated webpages:
    http://www.drunkmenworkhere.org/200
    http://memecodes.outer-court.com

17   *http://xroads.virginia.edu/~HYPER/POE/purloine.html*

18   Lacan, J. : "Le Séminaire sur la 'Lettre Volée'", *Ecrits*, Paris,
    Ed du Seuil, 1966

19   Lacan, J. : "Le temps logique et l'assertion de certitude anticipée",
    Ecrits, Paris, Ed du Seuil, 1966

20   Five discs are shown to three prisoners. Two are black and three white.
    One fixes one of these discs between the shoulders of each one, without
    those being able to know the color of it. Each prisoner can see the disc
    of both others, but not his own. There is no possible communication
    between them. The head warden proposes to release the prisoner who
    will logically establish the color of his own disc.

21   *http://www.cs.swarthmore.edu/~dylan/Turing.html*

22   For instance: *http://pespmc1.vub.ac.be/PRISDIL.html*
    http://plato.stanford.edu/entries/prisoner-dilemma