Logique & Analyse **218** (2012), 185–203

WHY WHITE HORSES ARE NOT HORSES AND OTHER CHINESE PUZZLES...

THIERRY LUCAS*

To the memory of J. Ladrière who began with logic and whose interests ranged over many fields of philosophy

The aim of this paper is on the one hand to remind the Western reader of some aporias of Chinese antiquity, and on the other hand to show that a logic of sorts or of types similar to that which has been proposed to explain the relation between categories (in the mathematical sense of the term) and logic brings much light on these aporias. This should be contrasted with older traditional explanations using conventional syllogistics or feeling satisfied with too simple explanations such as the confusion between inclusion and identity. The article is based on preceding papers of mine (see [LUC], [LUCa], [LUCb]), but stresses the basic unity of the solutions which I proposed there, a unity which is probably not apparent to the casual reader and which is shown here by sketching a very simple formal system and its semantics. I apologize for overlappings with some of my previous publications, but it seemed to me that the present paper would be unreadable if I just presented the final part without repeating the basic motivations.

A few words about the historical context

Between 500 BC and the accession to the throne of the famous emperor Qin Shi Huang Di, many schools of thought flourished in China. These schools evolved in what is now called Confucianism, Daoism, Moism, Legalism, Ming Jia (School of Names) and produced texts on the military art, on mathematics, etc. The historical context is well known but should be much more developed in a paper less exclusively oriented towards logic (see e.g. [FEN], [GRA]) as this one. But we should say a few words on how logical discussions arose at that time. After the Xia and Shang dynasties, whose

*Emeritus Professor at Université catholique de Louvain.

history comes to us tainted with myth, the house of Zhou accessed the imperial throne in 1045 BC, but began to lose its influence during the so-called Spring and Autumn period (722–476 BC): the feudal system was gradually collapsing and it was a period of political unstability and of continuous wars between states, the Warring states period (476–221 BC). A critique of the institutions naturally took place and thinkers circulated from state to state. Some thinkers like Confucius would plead for a restoration of the Zhou order, but others like Mozi, probably educated as a craftsman, would propose a new social order. His socio-ethical preoccupations are obvious, but the necessity to defend his point of view forced him and his school to develop a preoccupation for argumentation and an attention to language, which are the clear marks of logic. The period even produced more provocative people like Gongsun Long and Hui Shi whom people of the time considered as sophists. We will restrict our attention here to Gongsun Long and Mozi.

Gongsun Long and his "Discourse on the white horse"

Gongsun Long (320–250 BC) is an author who is famous among Chinese philosophers for his provocative assertion that white horses are not horses. We have a few texts attributed to him but some of them seem not to be authentic. The question of authenticity will not occupy us here, because we will concentrate on two texts for which authenticity is not challenged. There are many comments on those texts and as has been emphasized by Fraser in the Stanford Encyclopedia of Philosophy (entry "School of Names", [FRAc]), it is not sure that we have to consider Gongsun Long's "argumentation" more seriously than Lewis Carroll's puzzles. We think however that those texts reveal much about their epoch and that their logical and sophistical structure is really worth being investigated.

We first turn to *Baimalun*, the Discourse on the white horse. To clarify the discussion, I follow here very closely Fraser's presentation of the arguments in Stanford Encyclopedia of Philosophy (entry "School of Names"). Fraser presents the arguments in a "pidgin English, omitting articles and plurals" which sticks to the Chinese original: the reader unfamiliar with Chinese should know that classical as well as modern Chinese has no article, almost no plural, no gender, no declension, no conjugation so that classical texts are particularly open to many different interpretations: the provocative assertion 'Bai ma fei ma' may be rendered in English by " 'a white horse is not a horse', 'white horses are not horses', 'a white horse is not an exemplar of the kind horse', or 'the kind white horse is not identical with the kind horse'. " (Fraser)

Here are five arguments distinguished by Fraser; I quote from [FRAc]:

Argument 1. 'Horse' is that by which we name the shape. 'White' is that by which we name the color. Naming the color is not naming the shape. So white horse is not horse.

Argument 2. If someone seeks a horse, then it's admissible to deliver a brown or a black horse. If someone seeks a white horse, then it's inadmissible to deliver a brown or a black horse. Suppose white horse were indeed horse. In that case, what the person seeks in those two cases would be one and the same. What he seeks being one and the same is the white one not being different from horse. If what he seeks is not different, then how is it that the brown or black horse are in the one case admissible and in the other inadmissible? Admissible and inadmissible, that they contradict each other is clear. So brown and black horses are one and the same in being able to answer to "having horse" but not to "having white horse." This confirms that white horse is not horse.

Argument 3. Horses indeed have color; thus there are white horses. Supposing horses had no color, and there were simply horses and that's all, how could we pick out the white horses? So white is not horse. White horse is horse combined with white. Is horse combined with white the same as horse? So I say: White horse is not horse.

Argument 4. "Since you take having white horse to be having horse, we can say having horse is having brown horse, is that admissible?" "Not admissible." "Taking having horse to be different from having brown horse, this is taking brown horse to be different from horse. Taking brown horse to be different from horse, this is taking brown horse to be not horse. Taking brown horse to be not horse while taking white horse to be having horse, this is flying things entering a pond, inner and outer coffins in different places. These are the most contradictory sayings and confused expressions in the world."

Argument 5. "White" does not fix what is white... As to "white horse", saying it fixes what is white. What fixes what is white is not white. "Horse" selects or excludes none of the colors, so brown or black horses can all answer. "White horse" selects some color and excludes others; brown and black horses are all excluded on the basis of color, and so only white horse alone can answer. Excluding none is not excluding some. Therefore white horse is not horse.

A first examination of these arguments allows us to class them into two categories which I present according to contemporary terminology:

(1) A sophism (say S) playing on the ambiguity of the Chinese formulations of "White horse is horse" and "White horse is not horse"; one meaning is "White horses are horses" (true) and "White horses are not horses" (false); another meaning is "White horses are the same (thing or class) as horses"

(false) and "White horses are not the same (thing or class) as horses" (true). In terms of elementary classes, if Y represents the class of white horses and H represents the class of horses, we have of course $Y \subseteq H$ and $Y \neq H$. In a first order language using W and H as unary predicates to represent the fact to be white and the fact to be a horse, we have the corresponding semantic assertions:
$\models \forall x(Wx \land Hx \to Hx)$ but $\not\models \forall x(Wx \land Hx \leftrightarrow Hx)$.
In my opinion, arguments 2 and 4 are purely based on that sophism.

(2) An argument (say A, to avoid confusion with Arguments 1 to 5) which the jargon of the times classified as the problem of the "hard" and "white" (*jianbai*), alluding to the impossibility to separate those two qualities in real objects: you can have a stone which is hard and white, but you will not have an object which is hard without having a color or an object which is white without having some hardness. Argument 1 is particularly clear in that respect: it is considered as "sophistry" by philosophers of the time criticizing Gongsun Long because it "separates" hard and white, it separates "white" and "horse", color and shape, something which common sense does not allow. If we want to consider that argument as something other than sophistry or play on words, we can do justice to it simply by using the notion of category in Ryle's sense [RYL] or as you can find it in Carnap (cf his famous example, "Caesar is a prime number" [CAR]): "white horse" and "horse" are predicates of different categories, exactly as "Roman emperor" and "prime number" are predicates of different categories: white horse is no more horse than roman emperor is prime number. Read with that distinction in mind, argument 1 clearly distinguishes the category "color-form" and the category "form" and states very explicitly that those categories are not identical. By what we can see that argument A is subtler than sophism S: strange as it is, it has the merit of coherence. Arguments 3 and 5 also use that "separation" at least in the sense that they use explicitly the word "color".

If we want to represent that argumentation in a logic of sorts, we will have to introduce three sorts, a sort f "form", a sort c "color" and a sort s "form-color". Let us collect what seems necessary from a semantic point of view. The objects pertaining to these sorts cannot be simultaneously in two different sorts; in other words, a model $\mathcal{M}$ will have to determine sets $D_f^{\mathcal{M}}$, $D_c^{\mathcal{M}}$, $D_s^{\mathcal{M}}$, representing forms, colors and form-colors, or more exactly,
$D_f^{\mathcal{M}}$ representing the objects seen under the angle of form, color being abstracted;
$D_c^{\mathcal{M}}$ representing the objects seen under the angle of color, form being abstracted;
$D_s^{\mathcal{M}}$ representing the objects seen under the angle of form and color;
we can even be more precise about that model, considering that there is a

domain $D^{\mathcal{M}}$ of objects, say "our usual objects" and that those objects are "tripled"; they are seen once under the angle of form, once under the angle of color and once under the angle of both form and color; technically, this amounts to set:
$D_f^{\mathcal{M}} = D^{\mathcal{M}} \times \{f\}$
and similarly for $D_c^{\mathcal{M}}$ et $D_s^{\mathcal{M}}$.

Inside these domains, we will have to distinguish predicates. Thus,
- for sort f, we will have predicates P_f, Q_f, ... which will be interpreted by subsets $P_f^{\mathcal{M}}$, $Q_f^{\mathcal{M}}$, ... representing for example horse-objects, ox-objects, etc. seen under the angle of form;
- for sort c, we will have predicates R_c, S_c, ... which will be interpreted by subsets $R_c^{\mathcal{M}}$, $S_c^{\mathcal{M}}$, ... representing for example white-objects, brown-objects, etc. seen under the angle of color;
- for sort s, we will have predicates T_s, U_s, ... which will be interpreted by subsets which will be interpreted by subsets $P_s^{\mathcal{M}}$, $Q_s^{\mathcal{M}}$, ... representing for example white-horse-objects, brown-horse-objects, white-ox-objects, etc.

In that context, we already do "semantic justice" to Gongsun Long's argument A: white horses constitute a subset $Y_s^{\mathcal{M}}$ of $D_s^{\mathcal{M}}$, horses constitute a subset $H_f^{\mathcal{M}}$ of $D_f^{\mathcal{M}}$; those two subsets are disjoint since the domains $D_s^{\mathcal{M}}$ and $D_f^{\mathcal{M}}$ are themselves disjoint. "White horse is [indeed] not horse", in the very strong sense that no white horse is a horse and no horse is a white horse.

Gongsun Long and his "Discourse on indices and things"

We now turn to another famous and even more puzzling text of Gongsun Long, a text entitled *Zhiwulun*, "On indices and things". Even for the translation of the title itself there is much room for disagreement, for *zhi* functions as a name, meaning "index finger of the hand" and as a verb, meaning "pointing at", "indicating". The philosophical interpreation of *zhi* is also extremely diverse, ranging from "universal" for Feng Youlan (see [FEN]), and "meaning" for Graham (see [GRA]), to a more recent "pointing at" for Reding (see [RED]). As a logician looking at the formal aspects of things, I would not like to commit myself to any definite interpretation, but I thought that "index" would be a good translation for *zhi*, in that it reproduces almost exactly the two fundamental meanings of *zhi*. Here is the text; I adopt Derk Bodde's English translation of Feng Youlan (see [FEN]), with the systematic replacement of *zhi* by "indices" and without committing myself to Feng's platonistic interpretation in terms of "universals". Let me say however that very different translations have been proposed, see e.g. the already quoted Graham and Reding.

There are no things that are not indices (1), but these indices are no indices.

If the world had no indices, things could not be things (2). If, there being no indices, the world had no things, could one speak of indices? (3) Indices are what do not exist in the world (4). Things are what do exist in the world (5). It is impossible to consider what does exist in the world to be what does not exist in the world. In the world there exist no indices, and things cannot be called indices. If they cannot be called indices, they are no indices.

They are no indices, (and yet it has been stated above that) there are no things that are not indices. That there are no indices in the world, and that things cannot be called indices, does not mean that there are no indices. It is not that there are no indices, because there are no things that are not indices.

There are no things that are not indices, but these indices are no indices. That there are no indices existing in the world, arises from the fact that all things have their own names, which are not themselves indices. When they are not indices, to call them indices, would be to take indices to mean also what are not indices. It would be impermissible to take what are not indices as indices.

Indices, moreover, are what are held in common in the world. There are no indices existing in the world, but things cannot be said to be without indices (7). That they cannot be said to be without indices, means that there are none that are not indices. That there are none that are not indices, means (we return to our opening statement) that there are no things that are not indices. There are no indices that are non-indices. Indices that share themselves in things are non-indices. Supposing there were no thing-indices (6) existing in the world, who would there be to speak directly about non-indices? If there were no things existing in the world, who would there be to speak directly about indices? If there were indices but no thing-indices in the world, who would there be to speak directly about non-indices? Who would there be to assert directly that there are no things that are not indices? Moreover how can indices, which certainly in themselves become non-indices, depend upon things, and so only be indices when they share themselves in these?

I have proposed elsewhere [LUC] to bring some elements of clarification to that truly puzzling text using a formal approach as before, without committing myself to a definite interpretation of the concepts recovered by "index" and "thing". I start from the uncontroversial interpretation that the text deals with things, with indices and, what has puzzled interpreters, by the sudden appearance in the text (6) of "things-indices". This suggests that we consider a set T of things and a set I of indices (these indices are our sorts),

as well as their product $T \times I$ representing the things-indices of the text. Elements of $T \times I$ are then pairs of the form $< t, i >$ which I venture to read "thing t considered with respect to i"; i may be given here various interpretations, platonistic universal, or "meaning", or perceptual "aspect" or even a nominalistic interpretation, such as word ("horse" or "white"), etc.: $< t, i >$ would accordingly be the thing t considered with respect to the universal "horseness", or the thing t considered with respect to the meaning of "to be a horse" (whatever it is), or the thing t considered under the aspect or form "horse", or the thing t considered with respect to the word "horse", etc. This gives rise to a sorted model $\mathcal{M}$ where the sorts are given by I and the domains $D_i^{\mathcal{M}}$ are given by the pairwise disjoint sets
$D_i^{\mathcal{M}} = T \times \{i\}$.

We also have to represent when a thing t considered with respect to index i is or is not i, say when the thing t considered with respect to horseness i is indeed a horse. This suggests that for each index i we introduce exactly one predicate P_i which is interpreted only in sort i as a subset $P_i^{\mathcal{M}}$ of $T \times \{i\}$. When $< t, i > \in P_i^{\mathcal{M}}$, we represent the idea that t in relation with index i satisfies indeed P_i, e.g. the thing t with respect to whiteness is indeed white. Finally, I assumed two conditions in [LUC]:
(a) for each $t \in T$, there is at least one $i \in I$ such that $< t, i > \in P_i^{\mathcal{M}}$,
(b) for each $i \in I$, there is at least one $t \in T$such that $< t, i > \in P_i^{\mathcal{M}}$.
Note that, in general, for every $t \in T$ there will be many $i \in I$ such that $< t, i > \in P_i^{\mathcal{M}}$: this thing t may be white or horse or whatever. In the same paper, I was also suggesting that for fixed t, the set of all $< t, i >$ such that $< t, i > \in P_i^{\mathcal{M}}$ nicely represents the idea of thing-index.

With that model in mind, we can make sense of at least some of the puzzling assertions of the text, world being understood as the set of $D_i^{\mathcal{M}} = T \times \{i\}$:
(a) All things are indices (see (1)): by assumption (a), every t appears as some $< t, i >$ satisfying the corresponding P_i
(b) Indices are not in the world (see e.g. (4)), but somehow they are (see (7)). They are in the sense that they never appear isolated in the domains but as the second element of in general many $< t, i >$ such that $< t, i > \in P_i^{\mathcal{M}}$. They are not in the sense that indices i do not appear isolated, but as second components of pairs $< t, i >$: by condition (b), indices are always indices of something.
(c) Things are in the world (see (5)), but somehow they are not (combine (1) with (4)). They are in the sense that they appear as the first component of in general many $< t, i >$ such that $< t, i > \in P_i^{\mathcal{M}}$. They are not in the sense that things t do not appear isolated, but as first components of pairs $< t, i >$: by condition (a), things are always "things of some index".
(d) To clarify (6) and the subsequent text, observe that if there were no things,

T would be empty, $T \times I$ would be empty and there would be no more entity $< t, i >$ we can speak of or apply the corresponding predicate P_i to. Similar remarks apply with the counterfactual "if there were no indices".

This model emphasizes the nature of Gongsun Long's universe (or at least my interpretation of it!): an extremely "separated" universe, in which all things t are considered separately with respect to a quality or aspect or index i. This was already present in the first model, where we "separated" the object t with respect to its shape or to its colour.

For further reference, note here that the model which is proposed here for the *Zhiwulun* and the model given for the *Baimalun* are particular cases of an underlying common model $\mathcal{M}$:
give a non-empty set I, the elements of which may be called sorts;
for each $i \in I$, give a non-empty set $D_i^{\mathcal{M}}$ subject to the condition that for different $i, j \in I$, $D_i^{\mathcal{M}} \cap D_j^{\mathcal{M}} = \emptyset$;
we have a set of predicates $P, Q, R, ...$ each one of which has to be interpreted in one and exactly one $D_i^{\mathcal{M}}$.

We will relax all these conditions later, for example the last condition, in asking that each predicate be interpreted in some (not necessarily exactly one nor necessarily all) D_i.

The Moist reasonings in the "Small pick"

A few words on the historical context. We go back in time to look at Mozi and at his Moist school. Mozi (470–391 BC) seems to be a craftsman who developed his thought in a spirit of critique of the existing social system. His ideas are clearly motivated by socio-ethical preoccupations: selection according to merit, impartial care (previously often translated as "universal love"), opposition to fatalism, etc. His ideas have been transmitted to us in a collection of manuscripts also called the *Mozi*. For our concern, it is important to note that argumentation plays an important role in Mozi's writings: analogical reasonings, chains of reasonings, examples given by the Sages of antiquity, etc. It is even more essential to note that Mozi was followed by disciples, now referred to as "the later Moist school", which is roughly contemporaneous of the Chinese School of Names and which developed a very conscious and already quite elaborate use of logic. The writings of that school are also part of the book *Mozi* under the names of Canons first part, Canons Second part, Explanations of the Canons first part, Explanations of the Canons second part, Big Pick (also translated by Major Illustrations by Feng), Small Pick (also translated by Minor Illustrations by Feng). The Canons appear as a list of definitions, and the Big Pick and the Small Pick exhibit a very conscientious use of reasonings, which by all commonly accepted criteria should be classified as logical. We turn to the most explicit use of reasonings of the Small Pick.

The Small Pick is well-known for the paradoxical assertion that "although robbers are men, to kill a robber is not to kill a man", which usual criteria declare invalid. Remember the context: they promoted universal love ("impartial care"), but also wanted to keep social order, so had to answer the objection that one should love robbers and certainly not punish them. In fact the reasoning comes to us in a bunch of 5 types of reasoning, which typically deal with the transition from "X is/is not Y" to "ZX is/is not ZY", where ZX most often represents the concatenation of the words Z and X. Each type is well documented by quite a number of examples, but we will only give one or two examples of each type. For the reader interested in a more detailed discussion of the text, we refer to [LUCb]. We use Graham's translation in [GRA].

Of the thing in general, there are cases where
[Type 1:] *something is so if the instanced is this thing* [going from "X is Y" to "ZX is ZY"];
[Type 2:] *or is not so though the instanced is this thing* [going from "X is Y" to "ZX is not ZY"];
[Type 3:] *or is so though the instanced is not this thing* [going from "X is not Y" to "ZX is ZY"];
[Type 4:] *or applies without exception in one case but not in the other* [going from "X is Y" to "every X is Y" or from "some X is Y" to "X is Y" or...];
[Type 5:] *or the instanced in one case is this and in the other is not* [examples of Z , X, Y, T, U for which we have "ZX is ZY" but "ZT is not ZU"].

Examples of type 1:
- *A white horse is a horse. To ride a white horse is to ride a horse.*
- *Jack is a person. To love Jack is to love people.*
Those examples set no problems and should be contrasted with Gong Sunlong's provocative "White horses are not horses". Almost any type of logic accepts that type of transition. For example, in usual first-order logic, the first type of reasoning could be semantically justified as:
$\forall x(Wx \rightarrow Hx) \models \forall x(Wx \wedge Rx \rightarrow Hx \wedge Rx)$.
where W is the predicate standing for "to be white horse", H is the predicate standing for "to be a horse" and R is the predicate standing for "is ridden". The second example has the same form if we use the usual trick of describing the constant "Jack" by a predicate J "to be Jack", P and L having the obvious meanings:
$\forall x(Jx \rightarrow Px) \models \forall x(Jx \wedge Lx \rightarrow Px \wedge Lx)$.

Examples of type 2:
- Her younger brother is a handsome man, but loving her younger brother is not loving handsome men.
- [...] although robbers are men [...], killing robbers is not killing men.
By contrast to examples of type 1 this is setting a real problem. The usual conclusions of course would be that loving her younger brother is loving handsome men and that killing robbers is killing people. The idea of distinguishing sorts can be used here again: distinguish two meanings for loving, one related with the family sphere, the other one related with love within the sexual sphere; similarly, for killing, distinguish killing in the social sense of crime and killing in the moral sense of just punishment (needless to say that what we are discussing here is logic and not even gross approximations of ethics). To explain in more detail this second example, we will thus distinguish two sorts, say s for the "social" meaning and m for the "moral" meaning. Semantically speaking, very much like in the *Baimalun*, we will consider a model $\mathcal{M}$ given by a domain $D^{\mathcal{M}}$ ("our usual objects") and duplicate it according to the sorts we consider:
$D_s^{\mathcal{M}} = D^{\mathcal{M}} \times \{s\}$
$D_m^{\mathcal{M}} = D^{\mathcal{M}} \times \{m\}$
As in the *Baimalun*, we would like to consider that we do have two incomparable predicates, "to kill in the moral sense" and "to kill in the social sense", and on the other hand to recognize that there are here connections between the premisses and the conclusion which invite us to allow at least some comparability between the two sorts. Slightly abstracting and disregarding the precise form of $D_s^{\mathcal{M}}$ and of $D_m^{\mathcal{M}}$ as products of $D^{\mathcal{M}}$ with $\{s\}$ and $\{m\}$ respectively, we could obtain in a fell swoop the nice result that in some sense, we make $\forall x(Rx \to Mx)$ true, but $\forall x(Rx \wedge Kx \to Mx \wedge Kx)$ false. Define a model here by giving two sorts s and m, two domains $D_s^{\mathcal{M}}$, $D_m^{\mathcal{M}}$ and a function $h_{m,s}$ from the first one to the second one (if a represents a moral object, $h_{m,s}^{\mathcal{M}}(a)$ would represent the "social object underlying" a; if a represents a robber, object of the moral sphere, $h_{m,s}(a)$ would thus represent the man underlying a; with the above forms $D_m^{\mathcal{M}} = D^{\mathcal{M}} \times \{m\}$ and $D_s^{\mathcal{M}} = D^{\mathcal{M}} \times \{s\}$, we would naturally define $h_{m,s}(< b, m >) =< b, s >)$ for $b \in D^{\mathcal{M}}$). It is then easy to define subsets $K_m^{\mathcal{M}}$ and $R_m^{\mathcal{M}}$ of $D_m^{\mathcal{M}}$, $K_s^{\mathcal{M}}$ and $M_s^{\mathcal{M}}$ of $D_s^{\mathcal{M}}$ and to give an object a of $D_m^{\mathcal{M}}$ as well as a function $h_{m,s}^{\mathcal{M}}$ from $D_m^{\mathcal{M}}$ to $D_s^{\mathcal{M}}$ in such a way that:
$h_{m,s}^{\mathcal{M}}[R_m^{\mathcal{M}}] \subseteq M_s^{\mathcal{M}}$ and
$a \in K_m^{\mathcal{M}}$ but
$h_{m,s}^{\mathcal{M}}(a) \notin K_s^{\mathcal{M}}$.
With definitions of satisfaction which will later be made more precise, this is enough to obtain the result that

$\forall x(R_m x \to M_s h_{m,s} x) \not\models \forall x(R_m x \wedge K_m x \to M_s h_{m,s} x \wedge K_s h_{m,s} x)$
Looking back at examples of type 1, the reader acquainted with logic will immediately see that we can recapture them by stipulating that "there is only one sort": $m = t$ and $h^{\mathcal{M}}_{m,s}$ is the identity on $D^{\mathcal{M}}_m$ (which is equal to $D^{\mathcal{M}}_s$). We will of course add that in that case, the reasoning discussed in those examples are valid.

Examples of type 3:
- [...] reading a book is not a book, but to like reading books is to like books.
- Being about to fall into a well is not falling into a well, but to stop some someone being about to fall into a well is to stop him falling into the well.
These examples are in no way paradoxical and they are often neglected in the literature. Take the first example to illustrate our approach. Let Rx stand for "x is an act of reading a book", Bx for "x is a book". We could dismiss this example as stating that the truth of a premiss of the form $\forall x(Rx \to \neg Bx)$ is compatible with the truth of $\forall x(Rx \to Lx) \to \forall x(Bx \to Lx)$.
I argued in [LUCb] that this solution misses an important point, which is the connection between R and B: after all, reading books is about books and being about to fall in a well is closely related to falling in a well. Here I also emphasize that we should try and find the common point behind Moist reasonings of the different types. I prefer to interpret the premiss as stating almost explicitly that we have two different sorts: the sort "act", say s, of which reading a book would be an example and the sort "object", say t of which a book would be an example. In a model $\mathcal{M}$, the relation between sort s and sort t will be given by a mapping $h^{\mathcal{M}}_{s,t}$ which to an object a of sort s associates an object $h^{\mathcal{M}}_{s,t}(a)$ of sort t; in our first example, the mapping is that which to the act of reading a book associates the book which is read; in our second example, the mapping is that which associates to the act which is about to happen the act itself. Semantically speaking, we would already much better explain the structure of that "reasoning" as stating that there is a model in which $s \neq t$, $\forall x_s(R_s x_s \to B_t h_{s,t} x_s)$ is true and $\forall x_s(R_s x_s \wedge L_s x_s \to B_t h_{s,t} x_s \wedge L_t h_{s,t} x_s)$ is also true.

Examples of type 4:
They will be omitted here, because in our opinion, they do not bring new insights about the "sort" interpretation.

Example of type 5:
If this horse's eyes are blind we say that this horse is blind; though this horse's eyes are big, we do not say that this horse is big.
This and similar examples are again examples of the type which has motivated people to introduce sorts. Considering the given example we will

interpret it by introducing two sorts, says s for "organ of an animal" (the eyes) and t for "animal" (the ox), and a function $h_{s,t}$ which to the organ x_s associates the animal $h_{s,t}x_s$ of which it is the organ: to the eyes, associate the ox they come from. Introduce also predicates C and D, interpreted in sorts s and t as follows: C_s as "to be blind as eyes", C_t as "to be blind as animal", D_s as "to be big as eyes", D_t as "to be big as animal". Semantically speaking, we would nicely explain the structure of that type of "reasoning" as stating that there is a model in which $s \neq t$ and for some C and D, we have that $\forall x_s(C_s x_s \rightarrow C_t h_{s,t} x_s)$ is true but $\forall x_s(D_s x_s \rightarrow D_t h_{s,t} x_s)$ is false. We refer the reader to [LUCb] for a more detailed discussion of that type of examples which we think give other elements supporting our "sort" interpretation of the Moist reasonings and gives a unified treatment of the Moist reasonings.

What is the logic of sorts which lies behind those examples?

Our aim in this paper is to exhibit a "minimal" logic of structured sorts which, according to our interpretation, underlies Gongsun Long and the examples of the Moist Canons. We do this here by proposing a language and a notion of model which has been sketched in different versions in our previous papers, but which could help the interested reader to compare with his favorite system. For our part, we will only mention two types of systems which are clearly "cousins" of ours. For simplicity and for lack of necessity here, we limit to 1 the arity of function symbols and of predicate symbols.

Allowing for the use of the same words interpreted in different sorts, we will associate with predicate symbols R a set $Def(R)$ of sorts on which it makes sense. By contrast, function symbols in those examples represent transitions between well identified sorts, so that to function symbols F, we will associate a definite couple of sorts, the first component of the couple representing the sort of the source and the second component representing the sort of the target. We give formal definitions, but refrain from entering into routine details or verifications.

A language L with sorts is given by
- a non empty set S of sorts;
- for each $s \in S$, a denumerably infinite set Var_s, the variables of type s;
- for each couple $< s, t > \in S \times S$, a set $\mathcal{F}_{s,t}$, the function symbols of type $< s, t >$;
- a set $\mathcal{R}$, the predicate symbols, together with a function Def associating with each $R \in \mathcal{R}$ a subset $Def(R)$ of S.

Terms of the language, their source and their target, are defined simultaneously by induction:

- if $x_s \in Var_s$, then x_s is a term of source s and target s;
- if T is a term of source s and target t, if F is a function symbol of type $< t, u >$, then FT is a term of source s and target u.

For formulas we require that atomic formulas respect types, but we admit formulas mixing types as in the following definition:
- if T is a term of target t, if $R \in \mathcal{R}$ and if $t \in Def(R)$, then RT is an (atomic) formula;
- if A is a formula, then $\neg A$ is a formula;
- if A and B are formulas, then $(A \wedge B)$, $(A \vee B)$, $(A \rightarrow B)$ are formulas;
- if A is a formula and x is a variable of any type, $\forall xA$ and $\exists xA$ are formulas.
Parentheses may be omitted according to usual conventions.

Going to semantics, we define interpretations of L.

A (semantic) interpretation $\mathcal{M}$ for language L is fixed as follows:
- for each sort $s \in S$, give a non empty set $D_s^{\mathcal{M}}$, the domain of objects of type s;
- for each $R \in \mathcal{R}$ and $s \in Def(R)$, give a subset $R_s^{\mathcal{M}}$ of $D_s^{\mathcal{M}}$;
- for each $F \in \mathcal{F}_{s,t}$, give a mapping $F_{s,t}^{\mathcal{M}}$ from $D_s^{\mathcal{M}}$ to $D_t^{\mathcal{M}}$.

We can define the notion of valuation v as usual, associating elements of the domains to variables but with the understanding that the association respects sorts: for $x \in Var_s$, $v(x) \in D_s^{\mathcal{M}}$.

Valuations of terms are defined by following their inductive definition; it will be proved that the value $v(T)$ of a term T of source s and type t associates to the element $v(x_s)$ of $D_s^{\mathcal{M}}$ the element $v(T)$ of $D_t^{\mathcal{M}}$; in less formal notation, $v(T(x_s)) = T^{\mathcal{M}}(v(x_s))$.

The notion of satisfaction of a formula A in a model $\mathcal{M}$ by a valuation v, $\mathcal{M} \models_v A$, is defined as in the classical case, the only notable modification being the case of atomic formulas:
- if A is an atomic formula of the form RT with a term T of target t and $t \in Def(R)$, then $\mathcal{M} \models_v A$ iff $v(T) \in R_t^{\mathcal{M}}$.

The notions of truth in a model, the notion of validity of a formula and of validity of a reasoning may also be easily adapted.

Syntax and semantics having been given, we may of course ask the traditional problems of axiomatizability and completeness, but we are not sure that they are worth the investigation without further motivation. We will therefore refrain from doing that here, but we think it interesting, (A) first to mention some possibilities of moderate extensions of the system, (B) secondly to explain how to relate this system with our preceding heuristic considerations on Gongsun Long and Mozi and (C) thirdly to sketch how to relate this system with other much more elaborate logical systems which have been proposed in the literature and are clearly cousins of the one which has been presented here.

(A) Some extensions of the system

Ad (A), an interesting addition is the consideration of equality "sort by sort", which would allow us to express the notions of injective or surjective mappings. Add a symbol $=$ to the symbols of our language, together with the
Formation rule for equality:
- if T and U are terms having the same target t, then $T = U$ is an atomic formula.
The corresponding semantic definition will be:
the valuation v satisfies $T = U$ in the model $\mathcal{M}$, in symbols $\mathcal{M} \models_v T = U$, iff $v(T) = v(U)$.
Injective and surjective $F^{\mathcal{M}}_{s,t}$ are then expressed by $\mathcal{M} \models \forall x_s \forall y_s (Fx_s = Fy_s \rightarrow x_s = y_s)$ and $\mathcal{M} \models \forall y_t \exists x_s (y_t = Fx_s)$ respectively. Usual assertions of finite cardinality are also at hand.

Still concerning (A), another quite natural extension which allows us to give a direct formalization of Mozi's second example of type 1, is the addition of constants; logically, these are zero-argument functions, for which we will refrain here from allowing an ambiguity of type. For such a consideration of constants, add a sort 1 to the set of sorts and allow for each $s \in S$, a set $\mathcal{F}_{1,s}$, the constants of type $< 1, s >$, together with the
Formation rule for constants:
- if a is a constant of type $< 1, t >$, then a is a term of source 1 and target s
In the definition of the (semantic) interpretation, one will first let $D^{\mathcal{M}}_1$ to be the 1-element set $\{0\}$ and ask that to each constant a of target t, one associate a mapping $a^{\mathcal{M}}_{1,t}$ from $D^{\mathcal{M}}_1$ to $D^{\mathcal{M}}_t$; this is of course "picking" one element in the set $D^{\mathcal{M}}_t$.

With constants and equality at hand and with a bit of theory of definition, we could now analyze Mozi's second example of type 1 more seriously by appealing to a formula like $\exists! x Jx$ which could be typed as $\exists! x_t Jx_t$.

To finish our remarks about (1), a more significant extension of the system would be to add binary predicates, sort by sort or more significantly mixing sorts (would "Caesar computed the 10000th prime number" be a good example?); the best way to handle such extensions would be to admit a coupling or product of types, a step which would be leading us to more elaborate systems which have been considered in the literature.

(B) How the system is related with our discussion of Gongsun Long and Mozi

Ad (B), it is clear that our minimal system is directly inspired by what we have said in our discussion of Mozi and the reader can check that the formulas and the models which have been used in that discussion agree with a minor difference which can be considered as purely notational: for $R \in \mathcal{R}$ and $s \in Def(R)$, write R_s; similarly, and for $F \in \mathcal{F}_{s,t}$, write $F_{s,t}$.

Gongsun Long's *Baimalun* is a very particular case of our system: $S = \{c, f, s\}$, $\mathcal{R} = \{W, H, Y\}$, $Def(W) = \{c\}$, $Def(H) = \{f\}$ and $Def(Y) = \{s\}$ and no function symbols. If we want to explain that Y is a combination of W and H, we could refine our approach by introducing a concatenation operation $*$ on predicate symbols and an operation $\times$ on sorts with adequate formation rules. But that is hardly worth the trouble, because we would have to interpret those operations in a wild way, without constraints, to avoid reintroducing a well behaved conjunction and a well behaved product of sorts. The case of *Zhiwulun* is similar.

We think that our approach nicely explains the similarities and differences between the two philosophers. Both are unconventional in their use of predicates, but Mozi keeps at least some relations between the different domains, while Gongsun Long disjoints our universe in many unrelated copies. We could perhaps say that Gongsun Long's universe is completely scattered, while Mozi's keeps some connections. Let us be more formal in this respect, by asking to Mozi the question: how do we distinguish "category-sensitive predicates" such as "is killed" and "category-insensitive" predicates such as "to be ridden". It is not enough to say that in the first case we need two different categories, while in the second case we need only one, for our counter-example heavily depends on the fact that in the model $\mathcal{M}$, $h^{\mathcal{M}}_{m,s}[K^{\mathcal{M}}_m] \not\subseteq K^{\mathcal{M}}_s$; more exactly, if that condition is not satisfied, there is no counter-example. This could lead to a formal definition such as:
Definition of insensitivity.

A predicate R is insensitive for categories $r, s \in Def(R)$ and function symbol $F \in \mathcal{F}_{r,s}$ within model $\mathcal{M}$ iff $F^{\mathcal{M}}_{r,s}[R^{\mathcal{M}}_r] \subseteq R^{\mathcal{M}}_s$,
a condition which in our language may also be written:
$\mathcal{M} \models \forall x_r(Rx_r \to RFx_r)$

We are of course expressing that F behaves homomorphically with respect to R. The definition may be extended to terms T of source r and target s (after proving a few elementary properties of terms) by replacing Fx_r by T.

We can also consider strengthening of that definition by quantifying over all sorts, or over all terms, or over all models.

These notions being given, we can imagine a modernized dialogue between Mozi and Gongsun Long:

- Gongsun Long: form is one sort f, color-and-form is another sort s and they have no relations with each other

- Mozi: My dear Long, I can accept with you to distinguish two sorts, the sort f and the sort s, but they do compare; to a given color-form-object, we can associate the underlying form-object; this gives us a mapping $F^{\mathcal{M}}_{s,f}$ from $D^{\mathcal{M}}_s$ to $D^{\mathcal{M}}_f$ and then you have to accept that something which is white horse when considered with respect to color and form is also horse when considered with respect to form: $\mathcal{M} \models \forall x_s(Yx_s \rightarrow HFx_s)$ (a).

- Gongsun Long: My dear Mo, I see you are forcing me to accept that a white horse is a horse; but how can you say that to ride a white horse is to ride a horse?

- Mozi: My dear Long, you will not maintain that the action of riding an animal depends on its color, wo'nt you? If so, you have to admit that the predicate R, "is ridden", is category insensitive in our interpretation $\mathcal{M}$ for the mapping $F^{\mathcal{M}}_{s,f}$, i.e. $\mathcal{M} \models \forall x_s(Rx_s \rightarrow RFx_s)$ (b). But you accept a minimum of logic, and from (a) and (b), you will also derive $\mathcal{M} \models \forall x_s(Yx_s \wedge Rx_s \rightarrow HFx_s \wedge RFx_s)$. But that means exactly that to ride a white horse is to ride a horse.

This conversation is a bit simplistic, but it shows how Mozi's examples of type 1 fit in his general scheme and shows that the main step consists in admitting a transition map from the sort "color and form" to the sort "form".

(C) How the system is related with other systems

People interested in the relation between logic and categories will also have noticed that our approach is quite closely related with the languages associated with mathematical categories, such as have been developed by W. Lawvere, J. Bénabou, B. Mitchell and others. It is easier to explain here how those languages arise when one wants to associate a language L (the internal language in the sense of Bénabou) to a mathematical category $\mathcal{C}$: to every object C of $\mathcal{C}$ associate a sort s_C and a set Var_C of variables of type s_C; to every mapping $F \in Hom(C, D)$ in the category, associate a function symbol F^L of type $< s_C, s_D >$; and to every subobject $R \in Sub_C$ of the category, associate a predicate symbol R^L of type s_C. Such types of languages, terms and formulas may be interpreted in categories having enough structure. The conceptual advantage of that kind of instrument is that formulas and terms of the language allow one to do "as if" objects of the category had elements even if they have none. A detailed comparison of how to relate that type of system and ours is beyond the scope of this paper.

A very interesting approach to Aristotle's syllogistic, taking into account the difficulties of "sorting" the predicates has been given in [LMR]. The approach clearly adresses problems similar to the ones discussed here, but it is mathematically much more elaborate and inspired by topos-theoretic considerations.

Another approach which has some similarities with the present one is that of modal logic and its possible worlds. It could perhaps be a formal candidate to explain intensional interpretations of the second type of Mozi's reasoning. The connection is roughly this: instead of distinguishing different sorts, assign the differences of interpretation to different possible worlds; translating our analysis of Mozi's example of type 2, consider s as a world of evaluation and consider t as the one and only one world related to s by the relation of accessibility of modal logic. With that translation at hand, our model closely corresponds to a modal model $\mathcal{M}$ in which
$\mathcal{M} \not\models \forall x(Rx \rightarrow NMx) \rightarrow \forall x(Rx \wedge Kx \rightarrow N(Mx \wedge Kx))$
with a *de re* interpretation of necessity, symbolized by N. No doubt this approach may be refined, but it could be argued that "to kill" is not an intensional verb and we regret that in that approach Mozi's examples of type 3 and 5 and Gongsun Long's "arguments" do not seem to fit in very naturally.

(D) Inserting the present approach in a broader philosophical context

In this paper we have concentrated on some technical aspects of the "Chinese puzzles", but for people who are interested in the philosophical presuppositions of the present kind of approach, we will refer the reader to the numerous papers devoted to Gongsun Long and Mozi. We should in particular mention Chung-ying Cheng's work on Gongsun Long, especially his early papers which suggested that Gongsun Long is Platonic in Zhiwulun, a feature which lends philosophical support to our approach using the logic of sorts. Let me quote in detail Cheng's text in his 2007 article [CHEa] , p. 548: "In my earlier works of 1969 [CHE] and 1970 [CHS], I have suggested that G[ong] S[un] L[ong] is Platonic in the sense that he is an abstract realist. I have considered white as indicating whiteness and horse as indicating horsehood as they are abstract qualities which are abstracted from real experiences and given a status of independent status apart from the concrete experience of the world. I may even think that the indeterminate (budingzhe) could be an abstraction from all known determinations."

Another major source of information is also given by the numerous articles published in the Journal of Chinese Philosophy on Chinese Logic, on Gongsun Long and on Mozi. See [JCP] 24(2), 1997 devoted to Chinese Logic,

[JCP] 34(4), 2007 with six papers on Gongsun Long and [JCP] 35(3), 2008 devoted to Moism.

We finally refer the reader to the Stanford Encyclopedia of Philosophy, articles [FRAa], [FRAb] and [FRAc], where the reader will find numerous references to papers devoted to the subjects evoked here.

We conclude this paper by stating our conviction that logical analysis and logical techniques can help to clarify philosophical issues.

UCL
Institut Supérieur de Philosophie
Collège Désiré Mercier
place du Cardinal Mercier 14
1348 Louvain-la-Neuve
E-mail: Thierry.Lucas@uclouvain.be

BIBLIOGRAPHY

[CAR] CARNAP Rudolf, "Ueberwindung der Metaphysik durch Logische Analyse der Sprache", *Erkenntnis* 2, 1932, translated by Arthur PAP as "The Elimination of Metaphysics Through Logical Analysis of Language", in "Logical Positivism", ed. A.J. Ayer, New York, Free Press, 1959.

[CHE] CHENG Chung-Ying, "Logic and Language in Chinese Thought", in *Contemporary Philosophy: A Survey* 3, 1969, ed. Raymond. Klibansky Florence: Institute Internationale di Philosophia, pp. 335–347.

[CHEa] CHENG Chung-Ying, "Reinterpreting Gongsun Longzi and Critical Comments on Other Interpretations", *Journal of Chinese Philosophy*, 2007, 34(4), pp. 537–60.

[CHS] CHENG Chung-Ying & SWAIN Richard H., "Logic and Ontology in the Chih Wu Lun of Kung-sun Lung-Tzu", *Philosophy East and West*, 1970, 20(2), pp. 137–54.

[FEN] FENG Youlan, *A History of Chinese Philosophy*, translated by Derk Bodde, Princeton, Princeton University Press, 1952.

[FRAa] FRASER Chris, "Mohist Canons" in *Stanford Encyclopedia of Philosophy*, first published 2005, revised 2009. http://plato.stanford.edu/entries/mohist-canons/

[FRAb] FRASER Chris, "Mohism" in *Stanford Encyclopedia of Philosophy*, first published 2002, revised 2010. http://plato.stanford.edu/entries/mohism/

[FRAc] FRASER Chris, "School of Names" in *Stanford Encyclopedia of Philosophy*, first published 2005, revised 2009. http://plato.stanford.edu/entries/school-names/

[JCP] *Journal of Chinese Philosophy*, edited by Chung-Ying Cheng, online ISSN: 1540-6253.

[LMR] LA PALME-REYES, MAC NAMARA J. & REYES G.E., "Functoriality and Grammatical Role in Syllogism", *Notre Dame Journal of Formal Logic*, vol. 35(1), 1994, pp. 41–66.

[LUC] LUCAS Thierry, "Hui Shih and Kong Sun Long: an approach from contemporary logic", *Journal of Chinese Philosophy*, 20(2), 1993, pp. 211–255.

[LUCa] LUCAS Thierry, "Later Mohist Logic, Lei, Classes and Sorts", *Journal of Chinese Philosophy*, vol. 32, no 3, 2005, pp. 349–365.

[LUCb] LUCAS Thierry, "The Logic of Mohist reasonings; Leis and structured sorts", *Universitas: Monthly Review of Philosophy and Culture* (Taipei, Taiwan, R.O.C.), vol. 37, no 8, 2010, pp. 65–93.

[GRA] GRAHAM A.C., *Later Mohist Logic, Ethics and Science*, The Chinese University Press, Hong Kong and School of Oriental and African Studies, London, 1978, 590 pp.

[RED] REDING Jean-Paul, "Gongsun Long on what is not: steps toward the deciphering of *Zhiwulun*", *Philosophy East and West*, 2002, 52(2), pp. 190–206.

[RYL] RYLE G., *The Concept of Mind*, The University of Chicago Press, Chicago, 1949.