School of Names and mental possibilities: Is a white horse a horse?

Miguel López-Astorga[1]

Abstract: According to the 'white horse' paradox, a horse that is white is not a horse: it is a white horse, but not a horse. Though logical arguments can be adduced in support of the proposition, people tend to reject the paradox. Individuals often regard a white horse as a horse and therefore usually ignore the arguments from logic. This paper attempts an explanation of their behaviour. The theory of mental models offers a cognitive explanation for the habitual rejection of the paradox. Within this framework, the key is the way that people reason – a way that does not necessarily coincide with logic.

Key words: logic, mental possibility, paradox, theory of mental models, white horse debate.

1. Introduction[2]

The 'white horse' paradox, habitually named 'when a white horse is not a horse' or the 'white horse dialogue' (e.g., Sun 2013), has been the subject of much discussion. It is habitually attributed to Gong Sun Long. The base argument seems to be in the first chapter of the book *Gong Sun Long Zi*, where a discussion between two people raising the paradox is described (see also, e.g., Indraccolo 2017). The paradox is related to the Chinese School of Names. This school goes back to the second century B.C.E. It reflected on the links between words and things (see also, e.g., Fraser 2017).

Much has been written about this paradox. Many works about it have been presented in both the East and the West, and from the most different perspectives (see, in addition, e.g., Graham 1990, Thompson 1995). Nevertheless, a consensual explanation of a fact related to the paradox might still be missing. Nowadays there is no

clear cognitive account of the reasons why average individuals without a logical background tend not to admit it. The essential sentence characterizing the paradox is (1).

(1)   A white horse is not a horse.

People often deem (1) as false. They think that a white horse is a horse. Nonetheless, (1) may not be a logical problem. Several explanations can be offered from the logical point of view, in which (1) seems to hold. So, it is necessary to explain why people tend to reject (1).

The present paper tries to give an answer to this problem. To do so, it resorts to the theory of mental models (e.g., Khemlani, Byrne & Johnson-Laird 2018). The paper is based on the way this theory describes information processing and language interpretation. As shown below, to take that way into account can lead to the reasons why individuals usually do not accept (1), and hence why they ignore any logical argument that can support it.

This paper has three parts. The first one develops an example of how (1) can be understood within the framework of logic. In fact, it will offer an argument in favor of (1) from that academic discipline. The second section is devoted to a short description of the theory of mental models. The key aspect of the theory to address will be the manner it moves away from logic. This aspect is important because it constitutes the basis of the final section. In the last section, it will be argued that people do not accept (1) because the human mind works in the way the theory of mental models describes. The point will be that the human mind does not always follow logic. Average people do not conform to logical laws and rules naturally and without knowledge of logic (e.g., Johnson-Laird 2010). Thereby, the reason why (1) is many times rejected can be that individuals interpret language by following processes such as those indicated by the theory of mental models.

2. The white horse paradox and state-descriptions

Actually, many accounts are possible for (1). These accounts may come not only from logic, but also from, for example, philosophy of language (this can be checked in the literature cited above). Some explanations have been especially relevant, for instance those based on the difference between the 'use' and 'mention' of words (e.g., Sun 2013, Thompson 1995). However, this section will present a very simple logical account. It is related to the distinction mentioned above and is not far from other explanations given in literature. It is based on the method of extension and intension proposed by Carnap (1947).

To use that method, it is necessary first to express (1) by means of logical forms. A possibility could be to take (2) and (3) to capture the concept of 'white horse'.

(2)   H

Where 'H' represents the fact of being a horse.

(3)   W

Where 'W' is the fact of being white.

Thereby, a formula such as (4) could express the concept of 'white horse'.

(4) $H \land W$

Where '$\land$' is a logical conjunction requiring the two conjuncts to be true.

With one more operator, that of equivalence, it would be possible to refer to the idea that 'a white horse is a horse'. That operator is generally represented by means of the symbol '$\leftrightarrow$'. It indicates that both its left formula and its right formula should be true or false at once. So, the logical equivalence can be used to establish definitions. (5) means that a white horse is a horse.

(5) $(H \land W) \leftrightarrow H$

An important point is that the logical equivalence can be considered an abbreviation of a more complex formula. Logical equivalence means that: a) there is a conditional relation from the left formula (which is taken as the antecedent) to the right formula (which is taken as the consequent), and b) there is also a conditional relation from the right formula (which is taken as the antecedent now) to the left formula (which is taken as the consequent now). In this way, (5) can be transformed into (6).

(6) $[(H \land W) \to H] \land [H \to (H \land W)]$

Where '$\to$' is the symbol for conditional relation.

The first conjunct in (6), that is (7), is not problematic.

(7) $(H \land W) \to H$

This last formula just means that the conjunction of being a horse (H) and being white (W) implies being a horse (H). The difficulties appear with the second conjunct in (6), that is (8).

(8)   H → (H ∧ W)

What (8) expresses is that the fact of being a horse (H) implies the facts of being a horse (H) and being white (W). Obviously, this may not be the case; a horse can have different colors. Hence, the fact of being a horse (H) does not imply the fact of being white (W).

Based on this, (5) is not true, and, accordingly, if taken as a definition, (1) is true. This is because it can be stated that (5) and (1) are contradictory. However, Carnap's (1947) approach can show all this even clearer.

An important concept in Carnap's (*ibid.*) framework is that of 'state-description'. State-descriptions are complete possible worlds. In them, all the atomic logical formulae (i.e., all the simple logical formulae without operators) that can be built are to be found, either affirmed or negated. State-descriptions are consistent: they cannot include a formula and its negation.

Another relevant concept of that very framework is that of L-equivalence. Following Carnap (*ibid.*), two formulae are L-equivalent if and only if they are true in exactly the same state-descriptions.

In this way, there is no doubt that there are state-descriptions in which, for example, H and W being deemed as predicates, H and W can be applied to the same animals (the state-descriptions in which all the horses are white). In those state-descriptions, (5) holds. Nevertheless, it cannot be said that the equivalence in (5) is a L-equivalence. It is evident that there are also state-descriptions with animals to which the predicate H can be applied but the predicate W cannot. This reveals that the state-descriptions in which (4) is true for a particular animal are not all the state-descriptions in which (2) is true for that very animal (i.e., this reveals that an animal is not white in all the state-descriptions in which that animal is a horse). Therefore, as said, there is no L-equivalence.

It is possible that arguments akin to those offered in this section have been already presented in literature. In fact, as mentioned, they are close to other previous accounts, and many works have been published about (1). At the same time, the explanation in this section is obvious: it can be derived from common knowledge. However, in any case, the aim of this section has been only to give a simple logical explanation of the reasons why (1) can be assumed. As it can be seen, that explanation can be raised from approaches of philosophy of language such as that of Carnap (1947), that is, that of his method of extension and intension.

3. Sentences and mental possibilities

However, there is a contemporary theory that, among other aims, tries to account for the real way people interpret language. That is the theory of mental models (see also, e.g., Ragni & Johnson-Laird 2020). It proposes a manner to process linguistic information that seems to move away from logic. As far as the present paper is concerned, it can be said that one of its main concepts is that of 'mental possibility' (see also, e.g., Byrne & Johnson-Laird 2020). According to it, connectives cause possibilities regarding to sentences to be displayed. The possibilities that can be deployed for a conditional such as (9) are evident.

(9)   If A, then B.

Following the literature (e.g., Byrne & Johnson-Laird 2020; López-Astorga & Torres-Bravo 2020), the three possibilities corresponding to (9) are the three main conjuncts in (10).

(10) Possible (A & B) & Possible (¬A & B) & Possible (¬A & ¬B)

Where '¬' symbolizes negation.

These three main conjuncts represent situations in which (9) can be true (see also, e.g., Johnson-Laird & Ragni 2019). Nevertheless, they may vary according to the content of the sentence (e.g., Quelhas, Johnson-Laird & Juhos 2010).
Disjunction can be another example. If inclusive, it would be similar to (11).

(11) Either A or B, or both of them.

In principle, the possibilities for (11) are those in (12) (see, e.g., Byrne & Johnson-Laird 2020, Johnson-Laird & Ragni 2019).

(12) Possible (A & B) & Possible (A & ¬B) & Possible (¬A & B)

Content can also change the possibilities in this case (e.g., Quelhas & Johnson-Laird 2017). Nevertheless, the interesting point here is what is a fact under the theory. When there are not several possibilities, as in (10) or (12), but only one possibility, that is a fact (see also, e.g., Khemlani, Hinterecker, & Johnson-Laird 2017). This happens, for instance, in the case of conjunctions such as (13). There is no 'set of possibilities' for them.

(13) A and B.

Conjunctions only express one possibility. That of (13) is that in (14).

(14) Possible (A & B)

But, given that in (14) there is only one possibility, it is not correct to name it 'possibility'. What (13) expresses is that A and B are facts (see also, e.g., Johnson-Laird & Ragni 2019).

On this basis, the theory of mental models can explain many intellectual behaviors. For example, it can account for why individuals can easily make inferences such as Modus Ponendo Ponens. This inference consists of two premises and a conclusion. The first premise is a conditional such as (9). The second one is a fact such as (15).

(15) A

And the conclusion is a fact such as (16).

(16) B

This inference is correct from the logical point of view. People usually have no problem in making it (e.g., Byrne & Johnson-Laird 2009). The reason for this is clear for the theory of mental models. (9) leads to the possibilities in (10). However, the second premise, that is, fact (15), removes all of the conjuncts of (10) in which A is not true, that is, the second and third conjuncts. The final result is just one possibility, that is, the first conjunct in (10), which matches with (14). Because it is only one possibility, as said, (14) refers to two facts: (15) and (16). This means that, if (9) and (15) hold, (16) holds too (an explanation similar to this one is to be found, e.g., in Byrne & Johnson-Laird 2009).

Nevertheless, the theory of mental models can also explain why human reasoning sometimes moves away from logic. In logic, it is valid to derive a sentence such as (11) from a fact such as (15). But it is not frequent that people make that inference (see, e.g., Orenes & Johnson-Laird 2012). Within the theory of mental models, the reason for this is obvious as well. (15) is a fact. Hence, it cannot allow possibilities in which it is denied. On the other hand, (11) is linked to the possibilities in (12). Therefore, the problem is that (15) is negated in the third conjunct in (12), which is (17).

(17) Possible (¬A & B)

Accordingly, if (15) is true, (11) cannot be admitted (an account akin to this one is to be found, e.g., in Orenes & Johnson-Laird 2012).

The theory of mental models deals with many more problems and results are presented in the literature about language and human reasoning. Nonetheless, what has been indicated can suffice for the aims of the present paper.

4. The theory of mental models and the white horse paradox

Based on all of this, the explanation from the theory of mental models of the reasons why people tend to reject (1) can be simple. As pointed out, the concept of 'white horse' can be divided into the concepts 'white' and 'horse'. So, it is evident that four possibilities can be thought of combining these two last concepts, mentioned in (18).

(18) Possible (white & horse) & Possible (white & not-horse) & Possible (not-white & horse) & Possible (not-white & not-horse)

Any statement containing the two concepts (as well as any connective linking them) would be a subset of (18). Thereby, if it is affirmed that there is a white horse, it is affirmed that the existence of a white horse is a fact. However, if the existence of a white horse is a fact, it is necessary to eliminate all the conjuncts in (18) in which one of the two concepts ('white' or 'horse') is not real. This implies ignoring all the conjuncts in (18) except the first one. The result would be (19).

(19) Possible (white & horse)

Indeed, (19) is a fact. As indicated, when there is only one possibility, that possibility is a fact. Nevertheless, what is most important here is that the statement that there is a white horse leads to consider just (19). The other three possibilities cannot be accepted if it is the case that there is a white horse. Accordingly, from the statement that there is a white horse, only a scenario akin to that described in (19) can be thought of. In that scenario, the existence of a horse is a fact. In this way, if there is a white horse, there is necessarily a horse. This can be the process explaining why, following the theory of mental models, people often deem (1) as a wrong sentence.

5. Conclusions

There is no doubt that more argumentations (different from that presented above) can be offered to support (1). These argumentations can come from logic, philosophy of language, or other academic disciplines. Likewise, frameworks on cognition and language other than the theory of mental models can be considered. Perhaps such frameworks could explain the reasons causing people to reject (1) too.

However, this paper has made a clear point. Arguments in favor of (1) are possible. Nevertheless, there are also contemporary theories with the potential to account for why (1) is not admitted in most occasions. One of those theories is that of mental models.

The literature can lead one to think that approaches such as that of Carnap (1947) and that of the theory of mental models are not so different. Carnap's (*ibid.*) proposal is related to modal logic. On the other hand, at least broadly speaking, modal logic has been linked to the theory of mental models as well (e.g., López-Astorga 2020). So, the connection can be conceived. But the aim of this paper was not to keep analyzing these possible relations but to offer an explanation of the reasons why individuals often reject paradoxes such as that in (1). This appeared to be necessary, since, as indicated, (1) can be deemed as a valid sentence within some frameworks (e.g., that of Carnap 1947).

Hence, an important conclusion can be drawn from all of this. That conclusion is similar to those of other works that have also used the theory of mental models to analyze other paradoxes or sophisms (e.g., López-Astorga 2016): the theory of mental models is able to explain phenomena as complex as those corresponding to (1). Therefore, it may be worth continuing to explore what this theory can offer by its way to understand the human mind and language.

The literature reveals several facts. The theory of mental models can account for why people often make some inferences successfully. In the same way, it can also describe the deduction processes that lead individuals to wrong conclusions (from the logical perspective). Examples in these two directions can be found in most of the references about the theory of mental models cited here. Thus, an instance of logical inference that people usually make correctly is Modus Ponendo Ponens. This has been addressed above by means of (9), (15) and (16). Similarly, an example of logical inference difficult to make is that related to disjunction also reviewed in this paper, that is, that related to the derivation of (11) from (15). However, the arguments of this work impinge to a greater extent on the linguistic dimension that the theory also has. From this point of view, the present paper is one more illustration of the potential of the theory of mental models in different areas of knowledge.

References

Byrne, R. M. J., Johnson-Laird, P. N. (2009), "*If* and the problems of conditional reasoning", *Trends in Cognitive Science*, 13/7, p. 282-287, https://doi.org/10.1016/j.tics.2009.04.003

Byrne, R. M. J., Johnson-Laird, P. N. (2020), "*If* and *or*: Real and counterfactual possibilities in their truth and probability", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46/4, p. 760-780, https://doi.org/10.1037/xlm0000756

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

Fraser, C. (2017), "School of Names", in Zalta, E. N. (ed.), *The Standford Encyclopedia of Philosophy* (Spring 2017 Edition), https://plato.stanford.edu/archives/spr2017/entries/school-names

Graham, A. C. (1990), *Studies in Chinese Philosophy and Philosophical Literature*, State University of New York Press, Albany.

Indraccolo, L. (2017), "The 'white horse is not a horse' debate", *Philosophy Compass*, 12/10, https://doi.org/10.1111/phc3.12434

Johnson-Laird, P. N. (2010), "Against logical form", *Psychologica Belgica*, 50/3-4, p. 193-221, http://doi.org/10.5334/pb-50-3-4-193

Johnson-Laird, P. N., Ragni, M. (2019), "Possibilities as the foundation of reasoning", *Cognition*, 193, 10.1016/j.cognition.2019.04.019

Khemlani, S., Byrne, R. M. J., Johnson-Laird, P. N. (2018), "Facts and possibilities: A model-based theory of sentential reasoning", *Cognitive Science*, 42/6, p. 1887-1924, https://doi.org/10.1111/cogs.12634

Khemlani, S., Hinterecker, T., Johnson-Laird, P. N. (2017), "The provenance of modal inference", in Gunzelmann, G., Howes, A., Tenbrink, T., Davelaar, E. J. (eds), *Computational Foundations of Cognition*, Cognitive Science Society, Austin, p. 663-668.

López-Astorga, M. (2016), "Mental models, logical forms, and the horns sophism", *Cogency*, 8/1, p. 7-19.

López-Astorga, M. (2020), "The modal square of opposition and the mental models theory", *Epistemology & Philosophy of Science*, 57/3, p. 75-86, https//doi.org/10.5840/eps202057343

López-Astorga, M., Torres-Bravo, L. D. (2020), "Formal inferences and their relations to knowledge acquisition: Mental models and semantic links", *Principia*, 24/2, p. 325-331, https://doi.org/10.5007/1808-1711.2020v24n2p325

Orenes, I., Johnson-Laird, P. N. (2012), "Logic, models, and paradoxical inferences", *Mind & Language*, 27/4, p. 357-377, https://doi.org/10.1111/j.1468-0017.2012.01448.x

Quelhas, A. C., Johnson-Laird, P. N. (2017), "The modulation of disjunctive assertions", *The Quarterly Journal of Experimental Psychology*, 70/4, p. 703-717, https://doi.org/10.1080/17470218.2016.1154079

Quelhas, A. C., Johnson-Laird, P. N., Juhos, C. (2010), "The modulation of conditional assertions and its effects on reasoning" *Quarterly Journal of Experimental Psychology*, 63, p. 1716-1739, https://doi.org/10.1080/17470210903536902

Ragni, M., Johnson-Laird, P. N. (2020), "Reasoning about epistemic possibilities", *Acta Psychologica*, 208, https://doi.org/10.1016/j.actpsy.2020.103081

Sun, Z. (2013), "Cracking the white horse puzzle", *Journal of East-West Thought*, 3/3, p. 97-106.

Thompson, K. O. (1995). "When a 'white horse' is not a 'horse' ", *Philosophy East and West*, 45/4, p. 481-499, https://doi.org/10.2307/1399790