COGENCY Vol. 8, No. 1 (7-19), Winter 2016 ISSN 0718-8285

Mental models, logical forms, and the horns sophism

Modelos mentales, formas lógicas y el sofisma de los cuernos

Miguel López-Astorga

University of Talca, Talca, Chile
milopez@utalca.cl

Received: 18-12-2015. **Accepted**: 09-05-2016.

Abstract: At present, there is a very relevant theory explaining human reasoning. That is the mental models theory, a semantic approach that is proving that it can solve most of the cognitive problems that are to be found in the scientific literature. In this paper, I try to show how this theory can account for the reasons why an ancient sophism indicated by Diogenes Laërtius is usually rejected by people, and why it is a better alternative than the thesis that the human mind works by means of formal rules. Likewise, I argue that to assume the mental models theory does not necessarily means to accept the idea that human reasoning is not logical.

Keywords: Formal rules; mental models; reasoning; sophism; standard logic.

Resumen: En el presente, existe una teoría muy relevante que explica el razonamiento humano. Se trata de la teoría de los modelos mentales, un enfoque semántico que está demostrando que puede resolver la mayoría de los problemas cognitivos que se pueden encontrar en la literatura científica. En este trabajo, trato de mostrar cómo esta teoría puede explicar las razones por las que un sofisma antiguo señalado por Diógenes Laercio es generalmente rechazado por los individuos, y por qué su marco se puede considerar una alternativa mejor que la tesis de que la mente humana opera en función de reglas formales. Del mismo modo, argumento que asumir la teoría de los modelos mentales no significa necesariamente aceptar la idea de que el razonamiento humano no es lógico.

Palabras clave: Reglas formales; modelos mentales; razonamiento; sofisma; lógica estándar.

1. Introduction

Today, it is really hard to propose ideas such as that standard logic leads human reasoning or that the human inferential activity is based on formal systems such as that of Gentzen (1935). Indeed, there are many difficulties with ideas in this way. One of them, for example, is related to the problems of Wason's selection task (Wason, 1966, 1968), a very simple conditional reasoning task that individuals do not often execute correctly. Another case is that of *Modus Tollendo Tollens* (p -> q, ¬q / Ergo ¬p; where '->' stands for conditional relationship and '¬' means denial), one of the *ἀναπόδεικτοι* (*indemonstrables*) that Diogenes Laërtius attributes to Chrysippus of Soli (Diogenes Laërtius, *Vitae Philosophorum* 7, 80), which, although it is valid in Gentzen's calculus, however, people do not always apply (see, e.g., Byrne and Johnson-Laird, 2009; López-Astorga, 2013). And two more examples can be those of the disjunction introduction rule (p / Ergo p v q; where 'v' represents disjunction) and of the conditional introduction rule (q / Ergo p -> q), other two rules absolutely valid in Gentzen's system and standard logic and that individuals only use under particular circumstances (see, e.g., López-Astorga, 2014a; Orenes & Johnson-Laird, 2012).

Thus, facts such as these make it impossible to state that standard logic (or at least standard logic alone) can explain the conclusions drawn in inferences by human beings. Nevertheless, in this scenario, a very interesting alternative is that provided by the mental models theory (e.g., Byrne & Johnson-Laird, 2009; Johnson-Laird, 2004, 2006, 2010, 2012, 2015; Khemlani, Lotstein, Trafton, & Johnson-Laird, 2015; Khemlani, Orenes & Johnson-Laird, 2014; Orenes & Johnson-Laird, 2012). This is a semantic approach that claims that people ignore logical forms in their reasoning processes and that they only take the semantic possibilities corresponding to sentences into account. The relevance of this semantic framework is obvious at present, since it can solve most of the cognitive problems that, in the current days, reasoning research needs to address, including those mentioned above, i.e., those of Wason's selection task, *Modus Tollendo Tollens*, the disjunction introduction rule, and the conditional introduction rule (see, in this regard, e.g., papers such as that of Byrne & Johnson-

Laird, 2009; that of López-Astorga, 2015; or that of Orenes & Johnson-Laird, 2012). In this way, given that many of the cognitive difficulties that this theory can overcome cannot be, apparently, accounted for by other contemporary theories, and that its methodology based on analyses of semantic possibilities has been used even in studies in which all the theses of the theory are not accepted (e.g., López-Astorga, 2014b, 2015), it seems to be worth checking whether or not the scope of the mental models theory is also wide enough to explain problems coming from classical antiquity. True, in several ancient sources we can find controversial issues related to logic and cognition, and, undoubtedly, one of these issues is that referring to sophisms, i.e., to incorrect arguments used in order to persuade or convince. So, it can be very enlightening to analyze sophisms from the perspective of the mental models theory.

In *Vitae Philosophorum* 7, 186-187, Diogenes Laërtius presents some of such sophisms and he assigns them to Chrysippus, although he has doubts whether one of them really belongs to Eubulides. All of those sophisms have evidently wrong conclusions, and it can be thought that any naïve individual (without logical training) can easily note that they are not correct. However, the last one, which is the one that could have be proposed by Eubulides, seems to be specially interesting, since it can clearly show that the mental models theory provides a very good explanation for the mental processes why people detect the mistake in it. To argue that, indeed, this is so is the main goal of this paper. Thus, I will try to prove that this later theory has machinery enough to account for what happens when individuals face arguments such as that indicated.

For this purpose, obviously, I will analyze that sophism based on the mental models theory. However, at the same time, I will also refer to what can be said for it from standard logic, and this will allow me to finish with a discussion on the reasons why it appears to be better to assume that human reasoning follows semantic models than to suppose that it works by means of formal rules, and with a description of the advantages that can have to accept the former possibility. Nevertheless, firstly, it seems to be appropriate to begin with an exposition of the theses of the mental models theory that will have to be considered in this paper.

2. The mental models theory and quantified sentences

The mental models theory provides accounts for very different types of sentences, but, as it can be checked below, the part of the theory that is interesting here is that related to universally quantified sentences, that is, to sentences of the kind 'every P is Q', namely, to the sentences that were stood for by the letter 'A' by the medieval philosophers devoted to Aristotelian logic. As it is well known, this kind of sentences is expressed in first-order predicate logic in this way:

(x) (Px -> Qx)

Where the brackets quantify universally the element between them (in this case, 'x') and P and Q are predicates that can be attributed to x.

But, as said, the mental models theory does not focus on the logical form of sentences, but on their semantic possibilities. The problem is that some of those possibilities are hard to identify and sometimes people only detect the easier ones. These later possibilities are named 'canonical models' in the theory, and the former 'noncanonical models'. Thus, based on table 1 in Khemlani et al. (2015, p. 5), it can be stated that, given a sentence of type A, at first individuals only pay attention to, for example, this canonical model:

P Q
P Q
P Q

Khemlani et al. (2015) use other expressions and letters, but what is relevant is that, as it can be noted, this model refers to three possible scenarios, and, in all of them, x is both P and Q. Only after further reflection individuals can realize that there are other possibilities as well. In this way, the example of noncanonical model for A sentences provided by Khemlani et al. (2015) is as follows:

P Q
¬P Q
¬P ¬Q

Now, although x continues to be Q in the two first scenarios, it is not P in the second one. On the other hand, x is neither P nor Q in the third one. Of course, it is evident that the second and the third scenarios remain consistent with a sentence such as 'every P is Q', but what is important is that, in this case, the individual has noted that, given the sentence, there are also other possibilities different from that in which just P and Q are true for x.

However, as mentioned, the later model is only an example of noncanonical model. The possibilities that can be detected by individuals might vary according to different factors related to semantic or pragmatic aspects. For example, the exact meanings of P and Q can lead them to note that some of those combinations are not actually possible, and that other combinations not included in the set can be considered too. Precisely, the sophism that is going to be analyzed in this paper can show this in a clear way.

Nevertheless, before reviewing that sophism, maybe it is necessary to add that, for the mental models theory, P and Q are not symbols, but representations of reality, and that letters such as these can only be used to streamline. Likewise, it must also be mentioned that, in Khemlani et al.'s table 1, other kinds of quantified sentences are included too (for example, particular affirmative sentences – stood for by the letter 'I' by medieval logicians – and universal and particular negative sentences – stood for by the letters 'E' and 'O' respectively by those same logicians). Besides, large parts of the literature on the mental models theory focus on propositional sentences without quantification, i.e., on sentences that would be considered simply conditionals, conjunctions, disjunctions… in standard propositional calculus. Nonetheless, as it can be checked below, all of these kinds of sentences are not relevant for this paper, since the sophism to be analyzed is mainly based on a sentence of type A in Aristotelian logic, namely, on a universal affirmative sentence.

In the next section, I try to show that the mental models theory can easily describe the mental process why individuals can come to the conclusion that the sophism is not acceptable, and that, on the other hand, a similar explanation just based on standard logic is very hard to provide. The point is that, as it can be seen below, the few theses of the mental models theory commented on are enough to do that.

3. The horns sophism

As said, Diogenes Laërtius is who indicates the sophism, which is exactly in *Vitae Philosophorum* 7, 187. Diogenes literally says,

"εἰ τι οὐκ ἀπέβαλες, τοῦτ' ἔχεις· κέρατα δ' οὐκ ἀπέβαλες· κέρατ' ἄρ ἔχεις."

[If you did not reject anything, you have that; but you did not reject horns: So, you have horns].

This is obvious, on the one hand, that this is an inference with two premises ('if you did not reject anything, you have that' and 'you did not reject horns') and a conclusion ('you have horns'), and, on the other hand, that the first premise is a sentence of type A. Therefore, in first-order predicate calculus, the argument can be formally expressed, for example, in this way:

(x) (¬Rx -> Hx)
¬Rh

Ergo Hh

Where 'R' is a predicate meaning 'rejected by you', 'H' is another predicate meaning 'had by you', and 'h' is a constant meaning 'horns.'

Because this inference is correct in first-order predicate calculus, in principle it cannot be explained, from this calculus, which the problem of the argument is. Indeed, only two easy rules of standard first-order predicate logic are needed to draw (Hh) from [(x) (¬Rx -> Hx)] and (¬Rh). The first one is the universal quantifier elimination rule [(x) (Px) / Ergo Pa; where 'a' is any constant] and *Modus Ponendo Ponens* (p -> q, p / Ergo q), another *ἀναπόδεικτος* that, according to Diogenes Laërtius, was provided by Chrysippus of Soli (*Vitae Philosophorum* 7, 80). In this way, the derivation could be as follows:

[1] (x) (¬Rx -> Hx)	(premise)
[2] ¬Rh	(premise)
[3] ¬Rh -> Hh	($_{(x)}$E; 1)
[4] Hh	(MP; 2, 3)

Where '$_{(x)}$E' represents the universal quantifier elimination rule and 'MP' *Modus Ponendo Ponens*.

So, as said, standard logic cannot account for the reasons why this inference is a sophism and tends to be rejected. Nevertheless, as also mentioned, the mental models theory does be able to do that. It is evident that, given that the first premise is a sentence of type A, its canonical model can be this one:

¬R H
¬R H
¬R H

Because H is true in all of the cases of ¬R, it can be thought that, if only this model is taken into account, the horns sophism should be accepted. However, the fact that the conclusion is rare (it does not match reality, since human beings do not have horns) can lead individuals to make further cognitive effort and to think about other possibilities. Nonetheless, as noted, the noncanonical model proposed by Khemlani et al. (2015) indicated above is only an example, and the real noncanonical model considered by individuals depends on pragmatic and semantic factors, and, in particular, on the meaning of the predicates. In this way, it cannot be expected that, in this case, the noncanonical model is this one:

¬R H
R H
R ¬H

And this cannot be expected at least for two reasons. Firstly, the second combination (R – H) does not seem possible, since it is difficult to assume that something (for example, horns) is rejected and it is had at the same time. Secondly, it appears to be lacking one more possible combination: ¬R - ¬H. True, it is absolutely possible that I do not reject anything and, at the same time, I do not have that, and the reason is that, if I do not have anything before, I do not reject it. Or, in other words, to reject something is absolutely necessary to have that in advance. Therefore, it seems that the most appropriate noncanonical model for the first premise is the following:

¬R	H
R	¬H
¬R	¬H

In this way, the second premise does not allow concluding anything now, since the information that something has been rejected is liked to two possibilities: the first one, in which that is had, and the third one, in which that is not had. So, from the datum that you do not have rejected horns, it cannot be deduced that you have horns, because, although the first combination (¬R – H) describes a situation in which that is so, the third one (¬R - ¬H) provides a scenario in which, while you has not rejected horns, you do not have them.

Thus, it is clear that the mental models theory can make explicit the reasons why this sophism is unacceptable by people, reasons that, apparently, cannot be given by first-order predicate calculus. However, the later non-canonical model can lead one to think that the problem for standard logic is that [(x) (¬Rx -> Hx)] is not the actual logical form of the first premise, and that the real logical form of it needs to be recovered. Certainly, if the possibilities are (¬R – H), (R - ¬H), and (¬R - ¬H), it appears that the most adequate relationship between (R) and (H) is not (¬R -> H), but (H -> ¬R), which in turn would lead us to this quantified sentence:

(x) (Hx -> ¬Rx)

Thus, the idea would be that the sentence was not well expressed in natural language, in this case, ancient Greek, and that the semantic analysis of the mental models theory shows that what it really means is: 'if you have something, you have not rejected it.' In this way, it would be clear why the sophism cannot be admitted, since it is not possible to draw [Hh] from [(x) (Hx -> ¬Rx)] and [¬Rh] in standard first-order predicate calculus.

But the problem is that this later explanation is not actually a first-order predicate calculus explanation. It depends on the theses of the mental models theory. The next section reflects about this point and the reasons why to assume that human reasoning is semantic appears to be a better explanatory alternative than to claim that it is based on formal or syntactic rules.

4. Mental models and logical forms

Indeed, the syntactic explanation that the horns sophism cannot be accepted because the real logical form of its first premise is not [(x) (¬Rx -> Hx)], but [(x) (Hx -> ¬Rx)] is not really a syntactic account. To note that the later formula is the correct one, it is necessary an analysis of possibilities similar to those that the proponents of the mental models theory often make. This is not a new idea. There are several works (e.g., López-Astorga, 2014a, 2014b, 2015) in which the real logical forms of certain propositions are recovered by means of analyses of semantic possibilities, and in which it is clearly shown that such recovery processes are not possible without assuming, at least as methodological tools, basic ideas of the mental models theory. So, any syntactic theory claiming that, before reasoning, it is necessary to identify the real logical forms of sentences by means of analyses of possible combinations would be indebted to the mental models theory.

But another important point is that, if the consideration of semantic possibilities already enables to explain human reasoning, why do we need to detect the real logical forms? True, by means of both the previous arguments and those of López-Astorga (2014a, 2014b, 2015), it can be noted that the accounts of the mental models theory can exist alone and are autonomous and independent. It is possible, for example, to explain the problems of the horns sophism from just the approach of the mental models theory. Nevertheless, as shown, it is not possible to do that from just standard logic. As indicated, this later logic needs the results that can be achieved based on the framework of the mental models theory to recover logical forms. Therefore, it is evident that the mental models theory is possible without standard logic, but a reasoning theory based on standard logic is not possible without the mental models theory. It hence is also obvious that it is not absolutely necessary to identify logical forms to explain reasoning (on this issue, Johnson-Laird's, 2010, arguments are very enlightening as well).

In fact, today there is no actually a theory holding that the human mind follows standard logic. It is true that there are formal or syntactic theories, such as, for example, that of mental logic (e.g., Braine & O'Brien, 1998; O'Brien, 2009, 2014; O'Brien & Li, 2013; O'Brien & Manfrinati, 2010). However, these theories do not assume standard calculus, but usually only

the rules that, following experimental results, people really appear to use. In addition, the formal theories do not seem to be able to explain all of the problems that the mental models theory can account for, including some of those commented on in the introduction of this paper. For instance, the mental logic theory can explain why people do not always use the disjunction introduction rule. The reason is that that rule is not a *Core Schema*, i.e., a schema used by individuals whenever it can be done. Nevertheless, it cannot explain why people do sometimes use this rule. On the contrary, the mental models theory does be able to account for when, why, and under what circumstances that same rule is applied. Following the mental models theory, the problem is that most of the time [p v q] cannot be derived from [p] because [p v q] admits the semantic possibility that [p] is false and [q] true. For example, 'the car is green or yellow' cannot be drawn from a sentence such as 'the car is green,' since the former admits the possibility that the car is not green and is yellow, and this is incompatible with the premise ('the car is green'). Nonetheless, if the premise is, for example, 'I bought food' and the conclusion is 'I bought food or pizza,' the inference does be acceptable, since it is not possible that the latter is true and the former false. And this is so because it is obvious that, if I buy pizza, I buy food, which means that it is not possible that I buy pizza and I do not buy food at the same time (on this issue, see, e.g., Orenes & Johnson-Laird, 2012).

Furthermore, the formulae of first-order predicate logic are very complex and sophisticated, as they include symbols that are difficult to relate to the expressions in natural language (for example, the universal quantifier). So, it is hard to imagine how such expressions can be translated into those symbols (see, e.g., López-Astorga, 2014a).

Therefore, it seems that Ockham's razor or the *lex parsimoniae* compels us to assume that human reasoning is semantic and that the best way to describe it is the way proposed by the mental models theory. At least for now, it appears to be obvious that, while logical forms can be recovered, that action is not necessary. That is an additional task, since first of all the semantic possibilities must be taken into account, and once the semantic possibilities are considered, to identify logical forms becomes a superfluous and trivial exercise.

5. Conclusions

The previous arguments show that something is clear. The mental models theory can describe the mental processes that lead us to reject sophisms. Of course, as explained, a formal account can be rebuilt, but, as also argued, that account necessarily depends on the review of semantic combinations proposed by the mental models theory. Besides, it is difficult to explain how the sentences in natural language become formulae of standard logic.

However, an interesting point in this regard deserves to be commented on. The acceptation of the mental models theory does not imply the idea that human reasoning is not logical, or is not somehow related to logic. The precedent pages and other papers about how the mental models theory deals with not quantified sentences (e.g., López-Astorga, 2014b, 2015) reveal that, based, of course, on the theses of the mental models theory, it is possible to give logical forms consistent with the arguments and the accounts of this later theory. Obviously, this action is *a posteriori*, and by this I do not mean that the real mental processes in reasoning can be both semantic and syntactic. I only mean that, if we assume the mental models theory, that does not lead us to reject the idea that reasoning is linked to logic. The mental processes may not be related to logical forms, and what people do in a natural way when they reason may have nothing to do with such forms. Nevertheless, given that it is possible to resort to logical forms, although these are sophisticated, coherent with the explanations of the mental models theory and its accounts of sophisms such as that of horns, it can be said that human thought has some kind of relation to logic (papers such as those of López-Astorga, 2014b, 2015, also propose the existence of some type of link in this sense).

It is evident that the mental models theory describes what reasoning really does, and that our thought is beyond standard logic. Nonetheless, while it is true that, if the mental models theory is assumed, it is necessary to accept that reasoning is semantic and not formal as well, the fact that rebuildings of logical forms can be made from the semantic analyses of the mental models theory indicates that standard logic should not be totally ignored. Logic does not lead human reasoning, but it appears that the lat-

ter is not incoherent with the former. Therefore, it continues to be possible that logic has a prescriptive role in the human inferential activity, although it does not describe that activity. Undoubtedly, it appears that further research in this way is not irrelevant.

Works Cited

Braine, Martin D. & David P. O'Brien. *Mental Logic*. Mahwah: Lawrence Erlbaum Associates, Inc., Publishers, 1998.

Byrne, Ruth M. & Philip N. Johnson-Laird. "'If' and the problems of conditional reasoning." *Trends in Cognitive Science,* 13(7) (2009): 282-287.

Gentzen, Gerhard. "Untersuchungen über das logische Schließen I." *Mathematische Zeitschrift,* 39 (1935): 176-210.

Johnson-Laird, Philip N. "The history of the mental models." In Manktelow, K. and M. C. Chung (eds.), *Psychology and Reasoning: Theoretical and Historical Perspectives* (pp. 179-212). New York: Psychology Press, 2004.

Johnson-Laird, Philip N. *How We Reason*. Oxford: Oxford University Press, 2006.

Johnson-Laird, Philip N. "Against logical form." *Psychologica Belgica,* 5(3/4) (2010), 193-221.

Johnson-Laird, Philip N. "Inference with mental models." In Holyoak, K. & R. G. Morrison (eds.), *The Oxford Handbook of Thinking and Reasoning* (pp. 134-145). New York: Oxford University Press, 2012.

Johnson-Laird, Philip N. "How to improve thinking." In Wegerif, R., L. Li, and J. C. Kaufman (eds.), *The Routledge International Handbook of Research on Teaching Thinking* (pp. 80-91). Abingdon and New York: Routledge, 2015.

Khemlani, Sangeet, Max Lotstein, J. Gregory Trafton & Philip N. Johnson-Laird. "Immediate inferences from quantified assertions." *The Quarterly Journal of Experimental Psychology,* 68(10) (2015): 2073-2096.

Khemlani, Sangeet, Isabel Orenes & Philip N. Johnson-Laird. "The negation of conjunctions, conditionals, and disjunctions." *Acta Psychologica,* 151 (2014): 1-7.

López-Astorga, Miguel. "Are conditional and disjunction really comparable?" *Universum,* 28(2) (2013): 229-242.

López-Astorga, Miguel. "The conditional introduction rule and human reasoning: Findings from the mental models theory." *Cogency,* 6(2) (2014a): 29-37.

López-Astorga, Miguel. "Mental models and syntactic rules: A study of the relations between semantics and syntax in inferential processes." *Analele Universitatii din Craiova, Seria Filosofie,* 33(1) (2014b): 107-117.

López-Astorga, Miguel. "Evolved mechanisms versus underlying conditional relations." *Studies in Logic, Grammar and Rhetoric,* 40(53) (2015): 241-253.

O'Brien, David P. "Human reasoning includes a mental logic." *Behavioral and Brain Sciences,* 32 (2009): 96-97.

O'Brien, David P. "Conditionals and disjunctions in mental-logic theory: A response to Liu and Chou (2012) and to López-Astorga (2013)." *Universum,* 29 (2)(2014): 221-235.

O'Brien, David P. & Shujing, Li. "Mental logic theory: A paradigmatic case of empirical research on the language of thought and inferential role semantics." *Journal of Foreign Languages,* 36(6) (2013): 27-41.

O'Brien, David P. & Andrea Manfrinati. "The mental logic theory of conditional proposition." In Oaksford, M. and N. Chater (eds.), *Cognition and conditionals: Probability and Logic in Human Thinking* (pp. 39-54). Oxford: Oxford University Press, 2010.

Orenes, Isabel & Philip N. Johnson-Laird, P. N. "Logic, models, and paradoxical inferences." *Mind & Language,* 27(4) (2012): 357-377.

Wason, Peter. "Reasoning." In Foss, B. (comp.), *New Horizons in Psychology* (pp. 135-151). Harmondsworth (Middlesex): Penguin, 1966.

Wason, Peter. "Reasoning about a rule." *Quarterly Journal of Experimental Psychology,* 20 (1968): 273-281.