THE POSSIBILITIES OF DISJUNCTION IN THE MENTAL MODELS THEORY

LAS POSIBILIDADES DE LA DISYUNCIÓN EN LA TEORÍA DE LOS MODELOS MENTALES

Miguel López Astorga (Universidad de Talca, Chile) <u>milopez@utalca.cl</u>

> Recibido: 12/7/2017 Aprobado:15/12/2017

ABSTRACT

Baratgin and colleagues have questioned certain aspects of the mental models theory related to disjunction. It is truth that, from this last theory, the paper authored by Baratgin et al. (2015) has already been responded. However, I try to further develop that response here by insisting in two important points of the theory: the role that modulation plays in it and the clear differences between its framework and standard logic. In this way, my main aim is to support to a larger extent, by means of theoretical arguments based on its general approach, the idea that the objections presented in the mentioned paper do not really impact the mental models theory.

Keywords: disjunction; mental models; modulation

RESUMEN

Baratgin y sus colaboradores han cuestionado ciertos aspectos de la teoría de los modelos mentales relativos a la disyunción. Es cierto que, desde esta última teoría, el trabajo de Baratgin et al. (2015) ya ha sido respondido. No obstante, aquí se trata de desarrollar en mayor medida tal respuesta insistiendo en dos puntos importantes de la teoría: el rol que la modulación desempeña en su seno y las claras diferencias que existen entre su marco y la lógica estándar. En este sentido, el objetivo principal es apoyar aún más, por medio de argumentos teóricos basados en su enfoque general, la idea de que las objeciones presentadas en el trabajo mencionado no tienen un impacto real en la teoría de los modelos mentales.

Palabras clave: disyunción; modelos mentales; modulación

Introduction

A very important framework explaining the human intellectual activity is nowadays the mental models theory (from now on, MMT). As shown by the literature on it, which is very large (just some examples of works about this theory are Johnson-Laird, 2004, 2006, 2010, 2012, 2015; Johnson-Laird, Khemlani, & Goodwin, 2015a, 2015b; Khemlani, Orenes, & Johnson-Laird, 2012, 2014; Oakhill & Garnham, 1996; Orenes & Johnson-Laird, 2012; Ragni, Sonntag, & Johnson-Laird, 2016), MMT can properly explain many cognitive facts that other approaches cannot. However, Baratgin et al. (2015) have raised some criticisms with regard to the way it deals with disjunction. In their view, there are a number of theoretical problems related to the sentences with the word 'or' that are unclear in the theory and that, therefore, allow questioning its accounts and results. The truth is that, from MMT, a

response to those criticisms has already been given (Johnson-Laird et al., 2015b). Nevertheless, I think that it is worth continuing to develop that response, since that can enable to show more clearly to what extent the machinery of MMT is powerful.

In this way, in this paper, I will mainly focus on the aspects questioned by Baratgin et al. (2015) in order to show that they are not really a problem for MMT, and that it can be even thought that some of them do not make sense under its framework. Nonetheless, to do that, I will firstly describe the general theses of MMT to which it is important to pay attention to achieve that aim. Then I will point out which the theoretical problems seen by Baratgin et al. are exactly. Finally I will argue that it is not hard to MMT to solve those difficulties and that some of them do not even exist in the theory.

MMT, Peirce, and disjunctions

Obviously, the part of MMT that is more interesting here is that related to disjunction. However, it cannot be ignored that it has a very strong philosophical basis coming from authors such as Peirce (1931-1958). Indeed, the concept of iconicity is very relevant for MMT (e.g., Johnson-Laird, 2012, p. 136) because one of its main ideas is that people do not usually relate sentences in natural language to logical forms or well-formed formulae in standard logic (e.g., Johnson-Laird, 2010), but to mental representations that reproduce different possibilities in reality. Those representations are named 'models', and the theory considers each connective joining clauses (e.g., the conditional, conjunction, disjunction,...) to have, in ideal circumstances, a set of 'Fully Explicit Models'.

By way of example, I will comment on only the Fully Explicit Models of disjunction, which, besides, it is necessary to know to understand and analyze Baratgin et al.'s (2015) criticisms. According to MMT, given an expression such as 'either p or q or both of them', in principle, we can think about three possibilities or models (see, e.g., Johnson-Laird, 2012, p. 138, Table 9.2):

[A]: p	q
--------	---

[B]: not-p q

[C]: p not-q

Evidently, an inclusive disjunction can be true when the two disjuncts are true [A], when the first disjunct is false and the second one is true [B], and when the first disjunct is true and the second one is false [C]. In this way, the only impossible situation is this one:

[D]: not-p not-q

And this is so because it is also evident that a disjunction can never be true when its two disjuncts are false.

Given the overlap with the cases in which disjunction is true or false in standard logic, it can be thought that MMT is akin to this last system. Nevertheless, really it is very far from that logic. On the one hand, [A], [B], [C], and [D] do not refer to cases in a truth table of standard logic. As said, they are iconic representations of the world describing the elements that can be possible in it. Thus, as I understand the theory, they all stand for the same world and the only difference between them is the presence or absence of 'p', 'q', or both of them.

But on the other hand, another relevant concept of the theory is that of modulation (e.g., Orenes & Johnson-Laird, 2012). Modulation refers to the fact that semantics and pragmatics can eliminate or change models. A clear example of its way of acting is given by this sentence:

Lucia watched the Barca or she watched football (Orenes & Johnson-Laird, 2012, p. 375).

This is a disjunction but its models set cannot be [A], [B], and [C], but only [A] and [B]:

[A]: Lucia watched the Barca she watched the football

[B]: Lucia did not watch the Barca she watched the football

[C] is not possible because the Barca is a football team, which means that is impossible to watch the Barca and not to watch the football.

There is no doubt that this is a very interesting aspect of MMT, since, in addition to show the importance that the meanings of sentences have in the human mind, it also reveals a clear difference between this theory and a natural deduction calculus similar to that of Gentzen (1934, 1935), and, therefore, standard logic. In this last logic, there is a rule, the disjunction introduction rule (from now on, DIR), which allows inferring a formula such as $[p \lor q]$ (where 'V' is disjunction) from a formula such as [p]. However, such a deduction is not valid in MMT, and the reason is that disjunction refers to a possibility that is incompatible with the premise. That possibility is, obviously, [C] because in it, as pointed out, 'p' is false, and that fact is inconsistent with the premise [p]. Orenes and Johnson-Laird (2012) checked in their experiments that, indeed, people tend to reject conclusions obtained by means of DIR. Nevertheless, they also showed another important strength of MMT: it is able to explain and predict which the exceptional cases in which DIR is accepted are. Such cases are those akin to the previous example, in which [C], the inconsistent possibility, is not. True, they also noted that, notwithstanding the above, most of their participants admitted as valid the conclusion in tasks similar to this one:

"Lucia watched the football. Does it follow that Lucia watched the Barca or she watched the football?" (Orenes & Johnson-Laird, 2012, p. 375).

The reason is, as explained, that, in cases such as that, the conclusion does not refer to a possibility incompatible with the premise (in this example, 'Lucia watched the football').

Those are the general lines of MMT that are relevant for the aims of this paper. Of course, the theory attributes, as indicated, Fully Explicit Models to other connectives as well, which can be, in the same way, modified, depending on the circumstances, by modulation. Nonetheless, as also mentioned, the criticisms provided by Baratgin et al. (2015) are focused on disjunction. So, what has been said can be considered enough to respond those criticisms.

Objections proposed by Baratgin and collaborators

Really, Baratgin et al. (2015) support an approach rival to MMT, the called 'New Paradigm'. This new paradigm, which also has an extensive literature (e.g., Adams, 1998; Adams & Levine, 1975; Oaksford & Chater, 2007, 2009; Pfeifer, 2012, 2015) is based on probability logic, and, although all of its proponents do not defend exactly the same theses, it seems that they share the idea that the valid inferences are those that have a conclusion with less uncertainty that the premises (e.g., Baratgin et al., 2015). However, it is beyond my goals here to review the main theses of this new approach and discuss about them. As said, my only aim is to analyze the objections against MMT that Baratgin and colleagues raise and show that this theory can face them. Actually, the paper authored by Baratgin et al. is a response to a previous work supporting MMT (Johnson-Laird et al., 2015a) and, as also indicated, some adherents of it have already responded in turn to Baratgin et al.'s criticisms (Johnson-Laird et al., 2015b). So, my arguments below will not ignore these two last works in favor of MMT and will try to develop some of their theses by means of other works about the theory cited above.

That said, it can be stated that one of the theoretical problems that Baratgin et al. (2015) see in MMT is that, according to it, a disjunction can only be true if none of its three possibilities, [A], [B], and [C] is impossible. This is a problem because, under this criterion, any contingent disjunction linking two disjuncts that have no necessary relationship of any kind between them (e.g., 'there is a cat or there is a dog') is true, and, nevertheless, a disjunction such as 'either p or not-p' is false. This last fact is so

because, while $[p v \neg p]$ (where '¬' is negation) is a tautology in standard logic, the possibility [A], which, in the case of the last sentence, refers to a scenario in which both 'p' and 'not-p' are true, is unacceptable.

Johnson-Laird et al. (2015b) respond to this objection stating that a disjunction whose two disjuncts are an element (e.g., 'p') and its denial (e.g., 'not-p') only has two models:

not-p

[C]: p

Note that really [B] represents here a situation in which 'not-p' is the first disjunct and 'not-p' is the second disjunct too, which means that in that scenario just 'not-p' happens. In the same way, [C] denotes a circumstance in which 'p' is the first disjunct and 'not-not-p', that is, 'p', is the second one as well, which leads to a representation in which just 'p' occurs. As far as [A] is concerned, "... yields the null model, akin to the empty set" (Johnson-Laird et al., 2015b, p. 549).

However, the machinery of MMT can provide more arguments. As indicated, modulation is an important part of the theory and, given that the meanings of 'p' and 'not-p' are incompatible, it can be argued that modulation removes the possibility [A] in this case. On the other hand, MMT does not consider, as standard logic, that disjunction is mainly inclusive. As explained, it also assigns Fully Explicit Models to other connectives, and one of them is precisely the exclusive disjunction (i.e., the kind of disjunction corresponding to sentences such as 'either p or q but not both'). Thus, in, for example, Johnson-Laird (2012, p. 138, Table 9.2), it is indicated that the combinations of possibilities of this type of disjunction are:

[B]: not-p q

[C]: p not-q

Obviously, from this point of view, it can be thought that a sentence such as 'either p or not-p' is clearly an exclusive disjunction that can be only related to [B] and [C], and that hence it only needs [B] and [C], and not [A], to be possible to be true.

Furthermore, the literature on MMT shows that, contrary to what seems to be involved by standard logic, if the contextual elements are not enlightening enough, people's most habitual tendency is to interpret disjunctions as exclusive. Indeed, the proponents of the theory have carried out experiments in this regard, and the results are very clear: the natural interpretation of a positive disjunction tends to be the exclusive one (e.g., Khemlani et al., 2014).

For all of these reasons, it appears that this first objection presented by Baratgin et al. (2015) is not a problem for MMT at all, since this theory has responses to it. Nonetheless, Baratgin et al.'s arguments continue and they claim that the theoretical position of MMT on DIR is also a problem. Their point is that, if following MMT, it is not correct to derive $[p \lor q]$ from q, it must be assumed that [q] and $[\neg(p \lor q)]$ are consistent, since, if a sentence cannot be inferred from another one, the denial of the former has to be consistent with the latter.

In this way, the difficulty is now that, for MMT, the models of a negated sentence are the opposite combinations to those of that same sentence in positive form (e.g., Khemlani et al., 2014). So, if the models of a disjunction such as $[p \lor q]$ are [A], [B], and [C], the only model of that very disjunction if negated is [D]. But [D], that is, a scenario in which both 'p' and 'q' are false, is incompatible with [q], which means that [q] and $[\neg(p \lor q)]$ cannot be consistent. In addition, based on previous literature, Baratgin et al. (2015) argue that people are often able to note that the probability of [q] should be equal to or lower than that of $[p \lor q]$, which suggests that most of the individuals can see what $[p \lor q]$ really means, and, therefore, as far as I understand their arguments, that, if [q] is true, $[p \lor q]$ must also be so.

However, MMT can respond to these objections in an easy way too. Firstly, an evident point in favor of it is that, as shown by Orenes and Johnson-Laird (2012), people really tend not to apply DIR, and this is a fact in which Johnson-Laird et al. (2015b) insist, as it seems that all the other cognitive approaches, including that of Baratgin et al. (2015), cannot account for it. As explained, according to MMT, 'either p or q' cannot be inferred from 'q' because the former enables the possibility [C], which is incompatible with 'q'. Obviously, there is no doubt that this alone is already a good account of what the experimental results often reveal, but, besides, allows developing additional arguments against Baratgin et al.'s (2015) paper.

That 'either p or q' cannot be deduced from 'q' in MMT does not mean that 'q' must be consistent with 'it is not the case that either p or q'. Indeed, in standard logic, it is absolutely true that, if a formula [x] cannot be derived from a formula [y], then the formula $[\neg x]$ has to be consistent with [y]. Nonetheless, an important aspect of MMT is that, as shown, it is not standard logic and, therefore, should not be assessed based on this last logic. The relationships between formulae that are valid in propositional calculus do not necessarily hold in MMT, which appears to be a theory that just describes what people usually do, and not a normative framework indicating what individuals should do. From this perspective, it can be claimed that MMT is not a logical calculus and that Baratgin et al. (2015) seem to consider it to be that.

MMT is, as mentioned, about iconic possibilities, and not about logical forms. In this way, given that, as also indicated, in the case of DIR, 'q' is compatible both with [A] and with [B], the only conclusion that can be drawn from the fact that people usually do not apply this rule is just that 'q' is compatible with the opposite of [C] (i.e., with the opposite of the only model that is incompatible with it), and not with the opposite of 'either p or q' in entirety (i.e., of [A], [B], and [C] altogether). Certainly, the opposite of [C] is a scenario in which 'p' is false and 'q' is true, and such a scenario is clearly consistent with 'q'. So, it can be thought that in MMT the impossibility to derive a sentence does not lead to assume consistency with the denial of that sentence, but simply with the denial of the model(s) that make(s) the derivation impossible.

Furthermore, the theory can explain the probabilities that people tend to attribute to sentences with the structures [p] and $[p \lor q]$. The former refers to one scenario and the latter to three possibilities. Hence it makes sense, from its point of view, that people think that the latter is more probable, because it is clearly more probable that one of three possibilities happens than that one possibility occurs. Thus, it can be stated that these objections raised by Baratgin et al. (2015) are not great problems for MMT either.

A last point is that Baratgin et al. (2015) also seem to claim that, given that a disjunction is true in MMT if its three models [A], [B], and [C] are possible, DIR should be always applied in the case of contingent disjunctions, which, as accounted for, can never be false, and it is clear that this does not happen. However, the reason of that is not hard to explain form MMT either. In DIR there is a premise, 'q', which causes one of the possibilities, [C], to be false, whether or not the content of the particular disjunction is contingent. Accordingly, in an inference with the structure of DIR, the disjunction is not true, since one of its models is not so.

Therefore, it seems that the arguments presented in Baratgin et al. (2015) against Johnson-Laird et al. (2015a) can be answered from the general approach of MMT without a lot of difficulties. Johnson-Laird et al. (2015b) already gave relevant responses in this regard. I have tried only to further develop such responses here.

Conclusions

This paper has addressed only a partial aspect of MMT, the one related to disjunction. As stated, the theory has a much broader scope, since, in addition to deal with other classical connectives habitually used in sentences, it has demonstrated that is able to explain the results in many experimental

reasoning tasks (see, e.g., all the literature on this theory cited above). Thus, MMT appears to have the potential to account for more phenomena than other alternative frameworks, as it solves problems impossible to explain for other approaches (e.g., the problems related to DIR).

Obviously, the theory is not perfect and still has details to face, as well as particular aspects that need to be reviewed. Nevertheless, this paper shows that such details and aspects do not refer to the objections presented by Baratgin et al. (2015). As argued above, these last objections are not obvious difficulties for MMT.

In this way, it can be thought that, given that, as mentioned, the literature on MMT reveals that it has an important empirical support, this framework can be used in different scientific fields such as, for example, philosophy, education, and linguistics. As also indicated above, the philosophical foundations of MMT are evident and, for this reason, it can become a very interesting tool of philosophical research. Nonetheless, it is even clearer that MMT can be of great help in activities and studies linked to the suitable interpretation of sentences, and the analysis of the real role that syntax, semantics, and pragmatics can play in communication.

Maybe it is only important not to forget two essential points of MMT to avoid criticisms such as those of Baratgin et al. (2015) and correctly use its machinery. On the one hand, the possibilities corresponding to sentences are not guaranteed, since modulation can modify them. On the other hand, certain properties such as consistency do not mean exactly the same in MMT as in standard logic, as the former is not the latter. MMT consistency is about particular models or possibilities, not about whole sentences or their logical forms. Probably, taking these points into account, it will be possible to continue to develop the theory and apply it in studies on different subject matters.

References

ADAMS, E. W. (1998). A Primer of Probability Logic. Stanford, CA: Center for the Study of Language and Information publications.

ADAMS, E. W. & LEVINE, H. P. (1975). On the uncertainties transmitted from premises to conclusions in deductive inferences. Synthese, 30, 429-460.

BARATGIN, J., DOUVEN, I., EVANS, J. St. B. T., OAKSFORD, M., OVER, D., POLITZER, G., & THOMPSON, V. (2015). "The new paradigm and mental models". *Trends in Cognitive Science*, *19*(10), 547-548.

GENTZEN, G. (1934). "Untersuchungen über das logische Schließen I". En Mathematische Zeitschrift, 39(2), 176-210.

--- (1935). "Untersuchungen über das logische Schließen II". En Mathematische Zeitschrift, 39(3), 405-431.

JOHNSON-LAIRD, P. N. (2004). "The history of the mental models". In MANKTELOW, K. & CHUNG M. C. (Eds.) *Psychology and Reasoning: Theoretical and Historical Perspectives.*. New York, NY: Psychology Press, 179-212.

JOHNSON-LAIRD, P. N. (2006). How We Reason. Oxford, UK: Oxford University Press.

--- (2010). "Against logical form". In Psychologica Belgica, 5(3/4), 193-221.

--- (2012). "Inference with mental models". In HOLYOAK, K. J. & R. G. MORRISON (Eds.). *The Oxford Handbook of Thinking and Reasoning*. New York, NY: Oxford University Press, pp. 134-145.

--- (2015). "How to improve thinking". In WEGERIF, R. LI, L. &. KAUFMAN J. C (Eds.) *The Routledge International Handbook of Research on Teaching Thinking*. Abingdon, UK, & New York, NY: Routledge, pp. 80-91.

JOHNSON-LAIRD, P. N., KHEMLANI, S., & GOODWIN, G. P. (2015a). Logic, probability, and human reasoning. *Trends in Cognitive Sciences*, 19(4), 201-214.

--- (2015b). Response to Baratgin et al.: Mental models integrate probability and deduction. *Trends in Cognitive Science*, 19(10), 548-549.

KHEMLANI, S., ORENES, I., & JOHNSON-LAIRD, P. N. (2012). Negation: A theory of its meaning, representation, and inference. *Psychological Review*, 109(4), 646-678.

--- (2014). The negation of conjunctions, conditionals, and disjunctions. Acta Psychologica, 151, 1-7.

OAKHILL, J. & GARNHAM, A. (Eds.) (1996). *Mental Models in Cognitive Science. Essays in Honour of Phil Johnson-Laird*. Hove, UK: Psychology Press.

OAKSFORD, M. & CHATER, N. (2007). Bayesian Rationality: The Probabilistic Approach to Human Reasoning. Oxford, UK: Oxford University Press.

--- (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences*, 32, 69-84.

ORENES, I. & JOHNSON-LAIRD, P. N. (2012). Logic, models, and paradoxical inferences. *Mind & Language*, 27(4), 357-377.

PEIRCE, C. S. (1931-1958). *Collected papers of Charles Sanders Peirce*. HARTSHORNE, C. WEISS, P. & BURKS, A. (Eds.). Cambridge, MA: Harvard University Press.

PFEIFER, N. (2012). Experiments on Aristotle's thesis: Towards an experimental philosophy of conditionals. *The Monist*, 95(2), 223-240.

--- (2015). The new psychology of reasoning: A mental probability logical perspective. Thinking & Reasoning, 19, 329-345.

RAGNI, M., SONNTAG, T., & JOHNSON-LAIRD, P. N. (2016). Spatial conditionals and illusory inferences. *Journal of Cognitive Psychology*, 28(3), 348-365.