

# REMINISCENCE THEORY AND IMMORTALITY OF THE SOUL: SOME MENTAL REPRESENTATIONS IN PLATO'S *PHAEDO*<sup>1</sup>

Miguel LÓPEZ-ASTORGA<sup>2</sup>

**Abstract:** *This paper focuses on a particular discussion Plato describes in *Phaedo*. That is the discussion between Socrates and Cebes about the pre-existence and immortality of the soul. The paper does not try to clarify or better understand the discussion. It is only intended to show that the mental representations underlying the arguments both Socrates and Cebes give are consistent with the possibilities the theory of mental models attributes to human reasoning. In this way, the present paper attempts to support the basic theses of that theory.*

**Keywords:** *mental models; mental representations; *Phaedo*; possibilities; soul.*

## Introduction

In *Phaedo*, Plato presents several arguments about the immortality of the soul. One of them is based on the idea that the soul pre-exists to the body: the soul lived in the world of ideas before. Accordingly, following Socrates, if the soul existed before the body, the soul should also exist after the body. However, Cebes gives reasons against this argument.

This paper will focus on this particular discussion between Socrates and Cebes in *Phaedo*. Nevertheless, its goal is not to illuminate to a greater extent that discussion. It is not to better account for the discussion either. The main aim of the present paper is to show that the arguments Socrates and Cebes use are related to mental representations compatible with the general theses of the theory of mental models (e.g., Khemlani, Byrne, & Johnson-Laird, 2018). This theory attributes to

---

<sup>1</sup> ORCID ID: <https://orcid.org/0000-0002-6004-0587>

## Acknowledgments

PIA Ciencias Cognitivas, Centro de Investigación en Ciencias Cognitivas, Instituto de Estudios Humanísticos, Universidad de Talca.

Fondo Fondecyt de Continuidad para Investigadores Senior, código FCSSEN2102, Universidad de Talca.

<sup>2</sup> Instituto de Estudios Humanísticos “Juan Ignacio Molina”, Universidad de Talca.

human reasoning certain mental possibilities or models (see also, e.g., López-Astorga, Ragni, & Johnson-Laird, 2021; Ragni & Johnson-Laird, 2020). Thus, what this paper will try to claim is that the mental representations linked to the arguments both Socrates and Cebes offer are very akin to the models the theory of mental models often assigns to human inferential processes.

Several arguments by ancient philosophers have been already considered in order to show that they are coherent with the aforementioned theory (e.g., López-Astorga, 2021). This paper is one more step in that direction. It keeps revealing that not only the data obtained in laboratories in experimental situations with participants can support the theory. The literature (e.g., López-Astorga, 2021) indicates that the analysis of philosophical arguments can bring out that the conclusions of those arguments are also consistent with the predictions of the theory of mental models.

To take that step, first the present paper will describe the particular argument given by Socrates and Cebes' reply to it in Plato's *Phaedo*. Then it will show some essential theses of the theory of mental models necessary to make its point. Finally, it will argue that the theory can capture the reasoning processes involved in the discussion between Socrates and Cebes.

### **Pre-existence and immortality in *Phaedo***

As it is well known, in *Phaedo* Socrates speaks about the pre-existence of the soul (from 76c on). According to Plato, knowing is actually reminding; things in this physical world remind us ideas we previously saw in another world. That world is the world of ideas, that is, the place where the soul lived before being incarnated in a body.

Based on this, Socrates' argument is that, if the soul existed before the body, then the soul is independent of the body. Hence the life of the soul does not depend on the life of the body. This means that the soul does not necessarily die when the body dies. What is most reasonable is to think that the soul continues to live after the death of the body.

But Cebes objects this argument (from 77c on). Cebes accepts the reminiscence theory, that is, that knowing is reminding and that, therefore, the soul pre-existed. Nonetheless, in Cebes' view, that does not prove that the soul will keep existing after death. That is not really a proof of the immortality of the soul.

In fact, Cebes offers an additional counterargument resorting to the example of a weaver (from 87a on). A weaver tailors a lot of clothes in life. If we think about all of those clothes ignoring the last ones, we can assume that the weaver always continues to exist after tailoring clothes and that, accordingly, the weaver is immortal. However, this argument is absurd. The last clothes cannot be ignored. There are final clothes. After them, the weaver cannot tailor clothes anymore because the weaver dies.

Likewise, the fact that the soul can survive the body once or more occasions does not imply that the soul is immortal either. As in the case of the weaver, it is possible that a body is the last one in which the soul is incarnated. After that body, the soul would die. So, Socrates' argument cannot be admitted.

The arguments given by both Socrates and Cebes can be translated into mental representations such as those the theory of mental models attribute to sentences and reasoning processes. That will be shown below. Before that, the next section describes the general characteristics of the theory.

### **The theory of mental models: models as iconic representations of reality**

Perhaps three points of the theory of mental models are important, at least as far as the aims of this paper are concerned: its modal character, its idea of iconicity, and its concept of modulation. The theory of mental models is a modal theory because it proposes that people, when processing information, think about different models in which that information can be true. Those models are possibilities, and each of those possibilities captures a situation in which the different propositions offering the information can be true or false (e.g., Khemlani, Hinterecker, & Johnson-Laird, 2017). This can be seen in the case of a disjunction such as (1).

- (1) She has a ball or a racket, or both of them.

If it is assumed that 'p' represents that she has a ball, and 'q' that she has a racket, following the theory of mental models, the possibilities or models corresponding to (1) would be those in (2) (see also, e.g., Johnson-Laird & Ragni, 2019).

(2) Possible ( $p \ \& \ q$ ) & Possible ( $p \ \& \ \neg q$ ) & Possible ( $\neg p \ \& \ q$ ).

In (2), ‘&’ works as the logical conjunction, that is, it expresses that the two conjuncts are true (see also, e.g., Khemlani et al., 2018). ‘ $\neg$ ’ is negation.

What (2) describes is a ‘conjunction of possibilities’. This is an important difference between the theory of mental models and classical logic. The possibilities in (2) seem to refer to rows in truth tables in which the inclusive disjunction is true. Nevertheless, this cannot be said. What links the possibilities is conjunction. So, the possibilities can be accepted at once. Models ( $p \ \& \ q$ ), ( $p \ \& \ \neg q$ ), and ( $\neg p \ \& \ q$ ) can be admitted at the same time because they are possibilities. Rows in truth tables cannot be assumed at once because, if one of them is the case, the other ones are false (e.g., Johnson-Laird & Ragni, 2019).

On the other hand, this account is not compatible with modal logic either. From (1), it can be derived that ‘she has a ball’ is possible:  $p$  is possible in the two first possibilities in (2). Nonetheless, that is an inference that is not usually correct in modal logic (e.g., Khemlani et al., 2017). Accordingly, the differences between the theory of mental models and logic are obvious.

But these are not the only differences. The possibilities in a conjunction of possibilities such as (2) are iconic (e.g., Johnson-Laird, Khemlani, & Goodwin, 2015). Each of them stands for reality as far as possible for the human mind. They are different alternatives of reality with little variations. For instance, the only difference between the two first possibilities in (2) is that she has a racket in the first one and not in the second one.

Finally, the possibilities in an inclusive disjunction are not always those in (2). They can be modulated. Semantics and pragmatics are very important in the theory of mental models: both of them can change the conjunctions of possibilities (e.g., Quelhas & Johnson-Laird, 2017). For example, one might think about a disjunction such as (3).

(3) You will go to Paris or, at least, you will go to France.

If now ‘ $p$ ’ refers to the fact that you go to Paris and ‘ $q$ ’ to the fact that you go to France, the conjunction of possibilities is transformed into (4).

(4) Possible ( $p \ \& \ q$ ) & Possible ( $\neg p \ \& \ q$ ).

The second possibility in (2) is missing in (4). That is because it is not possible to go to Paris without going to France (see also, e.g., Orenes & Johnson-Laird, 2012).

This is the case of disjunction. Perhaps that of the conditional is more interesting for the present paper. The situation is not very different in the case of this last connective. Given (5),

(5) If it is sunny, then it will be a great day.

If 'p' means that it is sunny and 'q' that it is a great day, the conjunction of possibilities is (6).

(6) Possible ( $p \ \& \ q$ ) & Possible ( $\neg p \ \& \ q$ ) & Possible ( $\neg p \ \& \ \neg q$ ).

However, a relevant point in this case is that the two last possibilities in (6) are presuppositions. If p is possible, as it is in the first possibility in (6),  $\neg p$  is possible too, as it is in the second and third possibilities in (6). But while p guarantees that q is true, with  $\neg p$ , q can be both true and false. That is the reason why the second and third possibilities are presuppositions (see, e.g., López-Astorga et al., 2021).

All it has been said for disjunction applies to the conditional as well. (6) is a conjunction of possibilities. Therefore, it does not represent rows in a truth table (see also, e.g., López-Astorga et al., 2021). From (5), it can be inferred, for instance, that 'it is sunny' is possible. This is because p is in the first possibility in (6). However, this is not often correct in modal logic (e.g., López-Astorga et al., 2021). On the other hand, the possibilities of a conditional are also iconic (see also, e.g., Johnson-Laird, 2012). In addition, semantics and pragmatics can work in the case of the conditional too (see also, e.g., Quelhas, Johnson-Laird, & Juhas, 2010). This last point can be seen with the following example:

(7) If the moon is a natural satellite, then you might like it.

If 'p' corresponds to the fact that the moon is a natural satellite and 'q' to the fact that you like it, its conjunction of possibilities is (8).

(8) Possible (p & q) & Possible (p &  $\neg$ q).

Conjunction of possibilities (8) is different from (6) for two reasons. The moon is in fact a natural satellite. Hence, all the cases of  $\neg$ p have been removed from the conjunction of possibilities. Regarding the consequent (q), the possibilities are two, since you may like the moon or not.

Johnson-Laird and Byrne (2002) distinguished ten interpretations of the conditional caused by the action of semantics and pragmatics. One of them is 'Conditional', which refers to (6). Another one is 'Strengthen Antecedent', which is equivalent to (8). There is one more interpretation interesting for the argumentation below. It is 'Ponens'; it is related to sentences such as (9).

(9) If they are teachers, then they teach.

If (9) is said about people that are teachers, by virtue of pragmatics, its conjunction of possibilities is not actually a conjunction of possibilities. The set includes just one possibility. If 'p' is that they are teachers and 'q' that they teach, the only possibility is the one in (10).

(10) Possible (p & q).

They are teachers. So, there are not any cases of  $\neg$ p. Because they are teachers, they must teach. Accordingly, there are not any cases of  $\neg$ q either.

Ponens is relevant in the theory of mental models. Following it, if, as in (10), there is only one possibility, that possibility is not really a possibility, but a fact (e.g., Khemlani et al., 2017). However, the importance of these interpretations will be seen below, where all of this is applied to the discussion between Socrates and Cebes in *Phaedo*.

### **The discussion between Socrates and Cebes: their mental representations**

It is usual to express as conditionals the relations in which a fact (or several facts) implies (or imply) another one. This is the case in classical logic. In this last logic, it is possible to build a conditional from the premises and the conclusion of an inference. The conjunction of the

premises can be considered as the antecedent of the conditional. The conclusion can be deemed as the consequent (see, e.g., Deaño, 1999). Furthermore, classical logic is not the only logic allowing building conditionals from inferences. For instance, it seems that Stoic logic enables that too (e.g., O’Toole & Jennings, 2004).

Given that the theory of mental models distances itself from logic, one might think that this does not occur in the theory. Nevertheless, that idea may not be correct. Without the action of semantics or pragmatics, the possibilities of a conditional are those in (6). Those possibilities reveal that, when the antecedent happens, the consequent necessarily happens as well. That is what the first possibility in (6) shows. A scenario in which the consequent is not the case is a scenario in which the antecedent also has to be false. That is what the third possibility in (6) indicates. Therefore, in principle and when there are not any influences from semantics or pragmatics, in the theory of mental models, the antecedents of conditionals lead to their consequents too. This means that the structure of the conditional can be suitable to present a proof, demonstration, or argumentation in this theory as well.

Thus, if the structure of the conditional is used to express the Socrates’ argument described above, (11) can capture it.

- (11) If the soul existed before the body, then the soul will exist after the body.

Because the soul was in the world of ideas before and without a body, one might think that the existence of the soul is independent from the body. In this way, (11) would well describe the argument. Thereby, if ‘p’ is linked to the fact that the soul pre-existed and ‘q’ to the fact that the soul exists after the body, the possibilities of (11) would be tentatively those in (6).

Nonetheless, Socrates assumes the reminiscence theory. So, in his view, there is no doubt that the soul existed before the body, that is, that p. This eliminates the possibilities with  $\neg p$  from (6), the final result being (10). But, if the result is (10), the result is a fact, since there is only one possibility in (10). And a fact cannot be challenged.

In this way, Socrates appears to understand (11) not as a conditional, that is, as a sentence whose possibilities are the ones in (6).

He seems to deem (11) as a case of Ponens, which reveals the fact indicated in the only possibility of (10). That fact is that the soul pre-existed (p) and that the soul has to keep existing (q).

However, Cebes' objection is that the correct relation between the pre-existence of the soul (p) and the subsequent existence of the soul (q) does not correspond to Ponens. According to Cebes, the correct interpretation is Strengthen Antecedent. Cebes accepts the reminiscence theory. So, he also rejects the mental representations in which  $\neg p$  appears. Those are the second and third possibilities in (6). Nevertheless, he does not only consider a fact such as that in (10). As the example of the weaver reveals, pre-existence does not ensure further existence. Hence, in addition to the possibility in (10), it is necessary to admit the second possibility in (8). This is because it is possible that the soul exists before the body and that, yet, after the body dies, the soul disappears.

Thereby, if Cebes' argument is that the possibilities are those in (8), that means that the discussion can be expressed by resorting to the mental representations both Socrates and Cebes consider possible. Socrates thinks that there is only one possibility, which matches (10): the soul pre-exists and also lives after life. Nonetheless, Cebes thinks that there is one more possibility: the soul pre-exists and does not continue to live after live. From his point of view, the pre-existence of the soul (p) does not assure the existence of the soul after death (q).

## Conclusions

As said, this is not the first time the theory of mental models is used to reconstruct the deep structure of an ancient philosophical argument or theory. In fact, the theory has been also taken to show that the interpretations of the conditionals employed by some Greek thinkers do not correspond to (6) either (e.g., López-Astorga, 2018). Nevertheless, this paper has focused on a particular case: an argument in favor of the immortality of the soul offered by Plato in *Phaedo* and the reply to it described in that very book.

According to Socrates, if the soul existed before the body, there is no reason to think that the soul will not keep existing after death. Nonetheless, Cebes raised that the fact that the soul previously exists does not imply the fact that the soul will subsequently exist. It is interesting that both of them agree with the reminiscence theory. Hence, both of them think that the soul lived before being incarnated in the

present body. The difference is in what they consider this fact to mean. Following Socrates, this fact necessarily provides the impossibility that the soul dies after the current life. So, there are not any mental representations or models that one can build to describe a scenario such as that one. For this reason, his interpretation is Ponens. However, Cebes admits that possibility. Thus, he has two mental representations that are transformed into a conjunction of two possibilities. The first conjunct or possibility is equivalent to the one proposed by Socrates: the one in which the soul pre-exists and continues to live after death. But the second conjunct or possibility reveals a situation incompatible with the first one: the one in which the soul keeps pre-existing but the soul dies before being incarnated in another body. Therefore, Cebes' interpretation is Strengthen Antecedent.

Psychological theories such as the theory of mental models are usually assessed from experiments with participants. Those participants run reasoning tasks. The results of those tasks are statistically analyzed. This allows checking whether or not the results are significantly consistent with the predictions of theories. There is no doubt that this activity is not only suitable, but also necessary. But, as shown in the literature (e.g., López-Astorga, 2021), to analyze arguments presented by thinkers throughout history can also be relevant in this sense. If those arguments are coherent with the predictions of psychological theories as well, that can be deemed as additional support to those theories. This is the task this paper has tried to do. The paper has resorted to an argument in favor of an idea and to a reply against that very argument in the same text. Both the argument and its reply have been evaluated under the theory of mental models.

## References

- DEAÑO, A. (1999). *Introducción a la lógica formal*. Madrid, Spain: Alianza Editorial.
- JOHNSON-Laird, P. N. (2012). Inference with mental models. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning* (pp. 134-145). New York, NY: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199734689.013.0009>
- JOHNSON-Laird, P. N. & BYRNE, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, 109(4), 646-678. <https://doi.org/10.1037//0033-295X.109.4.646>

- JOHNSON-Laird, P. N., KHEMLANI, S., & GOODWIN, G. P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences*, 19(4), 201-214. <https://doi.org/10.1016/j.tics.2015.02.006>
- JOHNSON-LAIRD, P. N. & RAGNI, M. (2019). Possibilities as the foundation of reasoning. *Cognition*, 193. <https://doi.org/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), 1887-1924. <https://doi.org/10.1111/cogs.12634>
- KHEMLANI, S., HINTERECKER, T., & JOHNSON-LAIRD, P. N. (2017). The provenance of modal inference. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Computational Foundations of Cognition* (pp. 663-668). Austin, TX: Cognitive Science Society.
- LÓPEZ-ASTORGA, M. (2018). Gorgias' argument does not include actual conditionals. *Problemas*, 93, 81-89.
- LÓPEZ-ASTORGA, M. (2021). *The Possibilities of the Mental Model Theory*. Curitiba, PR, Brazil: Appris Ltda.
- LÓPEZ-ASTORGA, M., Ragni, M., & Johnson-Laird, P. N. (2021). The probability of conditionals: A review. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-021-01938-5>
- ORENES, I. & JOHNSON-Laird, P. N. (2012). Logic, models, and paradoxical inferences. *Mind & Language*, 27(4), 357-377. DOI: <https://doi.org/10.1111/j.1468-0017.2012.01448.x>
- O'TOOLE, R. R. & JENNINGS, R. E. (2004). The Megarians and the Stoics. In D. M. Gabbay & J. Woods (Eds.), *Handbook of the History of Logic, Volume 1. Greek, Indian and Arabic Logic* (pp. 397-522). Amsterdam, The Netherlands: Elsevier.
- QUELHAS, A. C. & JOHNSON-LAIRD, P. N. (2017). The modulation of disjunctive assertions. *The Quarterly Journal of Experimental Psychology*, 70(4), 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, 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>