INDIVIDUAL DIFFERENCE IN DEDUCTIVE REASONING:

AS A FUNCTION OF WORKING MEMORY

AND

COGNITIVE ABILITIES

PING PING TSE

Supervisors

Sergio Moreno Ríos

Juan Antonio García Madruga

Facultad de Psicología

UNIVERSIDAD DE GRANADA

June 2017

Editor: Universidad de Granada. Tesis Doctorales
Autor: Tse Ping Ping
ISBN: 978-84-9163-461-4
URI: http://hdl.handle.net/10481/48188

Abstract of thesis entitled

Individual Difference in Deductive Reasoning:

As a Function of Working Memory and Cognitive Abilities

Submitted by

Ping Ping TSE

for the degree of Doctor of Psychology

at Universidad de Granada

in June 2017

Working memory (WM) and other cognitive abilities are found to be important for deductive reasoning performances. Four studies which examined individual differences in deductive reasoning as a function of working memory (and other cognitive abilities) were performed in both adults and preadolescents. In the first study with preadolescents, we found reliable effect of verbal working memory in reasoning performances (propositional and syllogistic reasoning). However, in the second study with more cognitive measures, fluid intelligence was the most important predictor of reasoning performances for the first grader but the working memory capacities (visuospatial, verbal and central executive) for the second graders. We focused on syllogistic reasoning for the two studies with adults. WM capacity was important for generation task but not evaluation task. This suggests that people might use different strategies according to the task demand. The effect of number of mental models (according to the mental model theory) of the syllogisms in the generation task and the use of atmosphere/matching heuristics, particularly in the evaluation task, were also confirmed. In the last study, the inhibitory control ability (as measured by Stroop task) was found to be an important factor for solving conflict syllogistic problems. However, significant correlation with WM measures was found for no-conflict but not conflict problems. It is possible that conflict problems are too difficult (all were 3-model problems) and solving them require WM capacities beyond the limitations of some participants (ceiling effect). The results suggest a dual-process explanation of syllogistic reasoning. These results are consistent with the proposal of mental model theory. (255 words)

INDIVIDUAL DIFFERENCE IN DEDUCTIVE REASONING:

AS A FUNCTION OF WORKING MEMORY

AND

COGNITIVE ABILITIES

by

Ping Ping TSE

A thesis submitted in partial fulfillment of the requirements for

the Degree of Doctor of Psychology

Universidad de Granada

June 2017

*General Abstract

El razonamiento es un proceso cognitivo complejo que requiere del empleo de recursos cognitivos. La memoria operativa (MO) y otras habilidades cognitivas han mostrado ser relevantes en la actuación deductiva. Este trabajo muestra cuatro estudios que examinan las diferencias individuales de preadolescentes y adultos en razonamiento deductivo, en función de la memoria operativa y de otras habilidades cognitivas. Los dos estudios con preadolescentes, además, investigaron factores de desarrollo en razonamiento. En el estudio uno, nos centramos en la memoria operativa verbal y en la capacidad de conocimiento metadeductivo, que es la base principal de los procesos empleados en el razonamiento Tipo 2. La preadolescencia es un periodo de desarrollo en el cual los chicos están completando la adquisición de ambas facultades. Se encontraron efectos significativos en la memoria operativa verbal en diferentes medidas de razonamiento (razonamiento proposicional y silogístico). Alternativamente, en el estudio 2, extendimos las medidas de capacidad cognitiva a inteligencia fluida (medida con el k-bit), control inhibitorio (tarea Stroop) y habilidad de comprensión, e incluimos, además, una medida de memoria operativa visuospacial (Corsi) y de ejecutivo central (n-Back), como predictores potenciales de la actuación en razonamiento de las diferentes tareas deductivas. Investigamos si, también, existe una interacción o efecto de desarrollo de las medidas. A diferencia del estudio 1, en el segundo, con más medidas cognitivas, la inteligencia fluida fue el predictor más importante en la actuación en las tareas de razonamiento (excepto en la tarea de generación de conclusión silogística) para los alumnos de primer curso, pero para alumnos de segundo curso el predictor fue la capacidad de memoria operativa (visuoespacial, verbal y ejecutivo central). El análisis de

regresión mostró que la inteligencia fluida fue el predictor más importante del razonamiento (excepto para la generación de conclusiones en la tarea de generación silogística) para los alumnos del primer curso, pero para los alumnos de segundo curso lo fueron las medidas de memoria operativa (medidas mediante las tareas de Reading Span, Corsi y nBak). Esto apoya la idea de que el desarrollo de la memoria operativa en la preadolescencia ayuda al desarrollo de las habilidades de razonamiento.

En los estudios con adultos nos centramos en analizar el razonamiento silogístico. En el estudio tres exploramos la importancia de la memoria operativa en las tareas de generación y de evaluación. La investigación en tareas silogísticas mostraron diferentes resultados para esos dos tipos de tareas. Por ejemplo, en la tarea de generación se encontró un mayor efecto figural de las premisas y un sesgo de creencia más potente que en la tarea de evaluación. La capacidad de memoria operativa fue relevante en la tarea de generación pero no en la tarea de evaluación. Esto sugiere que la gente podría hacer un uso diferente de las estrategias de acuerdo con las demandas de la tarea. El número de modelos mentales (de acuerdo con la teoría de los modelos mentales) en la tarea de generación y el uso del heurístico de emparejamiento/atmósfera, particularmente en la tarea de evaluación, también fueron confirmados. Junto a la memoria operativa, examinamos también la importancia del control inhibitorio en el razonamiento silogístico. Investigamos el uso de heurísticos y si los detalles semánticos de los silogismos son procesados o representados. Para examinar la importancia del control inhibitorio, en este estudio utilizamos un tipo de problemas de conflicto que requerían la operación de inhibición de heurísticos para lograr extraer conclusiones válidas. La habilidad para el control inhibitorio (medido mediante la tarea Stroop) se mostró como un factor importante al resolver los problemas silogísticos de conflicto. Es posible que los problemas de conflicto fueran tan difíciles (todos problemas de tres modelos) que su resolución requiriera de capacidades de memoria operativa que van más allá de las limitaciones de algunos de los participantes (por lo que ocurriría un efecto techo). Los resultados sugieren un explicación en términos de procesamiento dual del razonamiento silogístico. Estos resultados son consistentes con la propuesta de la teoría de los modelos mentales.

Los estudios mostraron la importancia de la memoria operativa (también de otras capacidades cognitivas y control inhibitorio) en diferentes tareas de razonamiento para ambos, preadolescentes y adultos. Los recursos de memoria operativa son esenciales para cualquier procesamiento que requiera esfuerzo. La limitación de estas capacidades, especialmente del ejecutivo central, siempre pueden dificultar la actuación en tareas complejas/complicadas, tales como las de razonamiento. Las tareas más difíciles, en términos de procesamiento y/o almacenaje, están más sujetas al error. En línea con la teoría de los modelos mentales y las teorías del procesamiento dual en las que la memoria operativa (en adultos) y el desarrollo de esta capacidad (en preadolescentes) es un factor importante para la ejecución en razonamiento, especialmente en problemas difíciles. Las diferencias en memoria operativa y en otras capacidades cognitivas es uno de los factores de las diferencias individuales en la ejecución del razonamiento, aparte de la motivación, los estilos de pensamiento y el uso de diferentes estrategias.

Declaration

I declare that this thesis represents my own work, except where due acknowledgement is made, and that it has not been previously included in a thesis, dissertation or report submitted to this University or to any other institution for a degree, diploma or other qualifications.

.

Signed

Ping Ping TSE

El doctorando / The *doctoral candidate* **[Tse PIng Ping]** y los directores de la tesis / and the thesis supervisors: **[Sergio Moreno Ríos y Juan García Madruga]**
Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

/

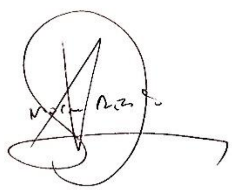
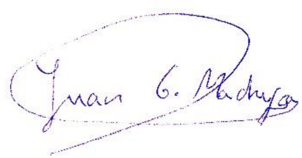
Guarantee, by signing this doctoral thesis, that the work has been done by the doctoral candidate under the direction of the thesis supervisors and, as far as our knowledge reaches, in the performance of the work, the rights of other authors to be cited (when their results or publications have been used) have been respected.

Lugar y fecha / Place and date:

Granada, 14-6-2017

Directores de la Tesis / *Thesis supervisors;* Doctorando / *Doctoral candidate*:

Firma / Signed Firma / Signed


Acknowledgements

I wish to express my heartfelt appreciation to my supervisors, Dr. Sergio Moreno Ríos, Prof. Juan Antonio García Madruga and Prof. María Teresa Bajo Molina, for their tremendous support and guidance throughout my studies. Their unfailing patience in reading and correcting my draft has been a great help to me, especially during the time when they were occupied with other works.

Special thanks to my colleagues who are always very supportive, especially to Olga and Vani, and to José Manuel Colino who spent hours helping me with the technical issues.

Thanks to my dearest and loveliest parents, brother and sister for their endless support. I love you always.

It would be certainly impossible to finish the thesis without their help.

Contents

'*' indicates that the chapter is written in Spanish

List of Tables and Figures

Chapter 1

Theories of deductive reasoning

1.1 Types of deductive reasoning

Syllogisms and propositional reasoning are most studied in the field of deductive reasoning. Classical syllogism consists of two premises with entities of three different categories (also known as "terms") connected with one of the four quantifiers: All, Some, Some . . . not, and None/No. These quantifiers are also known as Aristotle quantifiers. In logic, they act as the binder to denote the relationships between sets. In natural language, a quantifier is a determiner or pronoun indicative of quantity or amount. Quantifiers in syllogisms take their usual natural language meaning. For example:

Premise 1: Some As are Bs.

Premise 2: Some Bs are Cs.

Conclusion: (What, if anything, follows)

The presentation order of the three terms, As, Bs, and Cs, can differ. Different configurations of the terms are called figures[1], as follows:

---

[1] The notation by Johnson-Laird is used in this thesis. There is another notation by Aristotelians. Generally speaking, Figure 1, 2, 3 and 4 in Johnson-Laird notation are Figure 4, 1, 2 and 3, respectively, in Aristotelians. In additional, the entities are called "subject" (subject in the conclusion), "predicate" (predicate in the conclusion) and "middle term (appears in both premises but not in the conclusion) in Aristotle's definition.

	Figure 1	Figure 2	Figure 3	Figure 4
Premise 1:	A-B	B-A	A-B	B-A
Premise 2:	B-C	C-B	C-B	B-C

The four quantifiers options for each of the two premises and the four figures sum up to 64 possible syllogisms. The problems are abbreviated by letters according to the Latin derived names for the quantifiers -- All (A), Some (I), Some . . . not (O), and None (E) of each premise and the figure number, e.g., the example above is II1. The set relationship of A and C-terms is then referred from the two premises, assuming the two premises are true, to form a conclusion. In Johnson-Laird's definition, the conclusion can have two "directions": a-c or c-a. Either the A-term or the C-term can be the subject of the conclusion, while the other term being the predicate.

Three different inference tasks[2] are commonly employed in the studies of syllogisms, namely the generation, evaluation and verification tasks. In a generation task, participants are usually asked to write down/draw a conclusion which connect the A and C-terms with one of the four quantifiers, in either the a-c or c-a direction (e.g. All As are Cs or All Cs are As), or state that "no conclusion can be drawn". In the evaluation task, participants are asked to select the conclusion among certain number of options (usually 9[3]: the four quantifiers with the two different conclusion directions, plus "no valid conclusion"). Verification task in which participants have to decide whether the given conclusion is valid or not (two-alternative forced choice: valid or invalid) is also very common.

[2] Besides the inference tasks, modal task which asks about the necessity and possibility of the propositions is common also.

[3] We use Aac (all As are Cs), Aca (all Cs are As), Iac (some As are Cs), Ica (some Cs are As), Oac (some As are not Cs), Oca (some Cs are not As), Eac (no As are Cs), Eca (no Cs are As) and NVC (no valid conclusion) to denote the nine possible conclusions in this thesis.

Propositional reasoning includes conditionals (if, as known as implication), (inclusive) disjunction (or), exclusive or (p or q but not both), conjunction (and) and biconditional (if and only of). The “not” operator can be inserted in the arguments for negation. Among propositional reasoning, conditional reasoning is much more studied. Besides the inference tasks, other common tasks include the truth-table task, implication task and consistency judgement task. In nature language, a conditional argument is in the form “if p then q"” which expresses a causal relationship between the antecedent, p, and the consequent q, and similar for biconditional (which is in the form of “p if and only if q”). For the other three types, relationship between the two entities is connected by the operators (or, and, or but not both), such as p or q, p and q. In empirical studies, these arguments (as the first premise) are usually followed by an additional categorical information (“a fact”, as the second premise), i.e. p, q, not-p, or not-q, like the following two premises problem:

Premise 1 (conditional): If Lisa has an essay to write, she will study late in the library.

Premise 2 (fact): Lisa has an essay to write.

Almost all participants will draw the conclusion that “she will study late in the library” (known as modus ponens, that is to conclude the consequent from the conditional and given the antecedent as the fact). Other inference types include modus tollens, to conclude the negation of the antecedent (Lisa does not have an essay to write) from the conditional with the negation of the consequent as the given fact (i.e. she will not study late in the library). Both modus ponens (MP) and modus tollens (MT) are logically valid. While the other two inferences, denial of antecedent (DA), to conclude negation of the consequent when the negation of the antecedent is given as the additional fact and affirmation of consequent (AC), to conclude the antecedent when the consequent is given, are logically invalid but commonly drawn by humans.

Contrasting with inductive reasoning, deductive reasoning requires “top-down” processes to draw a conclusion by applying (reference) rules on one or more premises, with the assumption that the premise(s) is(are) true. The conclusion is necessarily true if the appropriate/correct rule(s) is/are applied. There are different theoretical views on the cognitive processes that underlie reasoning. The four major schools are the logic/rule-based theories[4]: e.g. the mental logic theory (e.g.,Braine, 1978; Braine & O'Brien, 1998; Rips, 1994), the mental models theory (e.g., Bucciarelli & Johnson-Laird, 1999; Johnson-Laird, 1989, 1999; Johnson-Laird & Byrne, 1991), the probabilistic approach (Chater & Oaksford, 1999; Oaksford & Chater, 1994; Oaksford, Chater, & Larkin, 2000; Oaksford, Roberts, & Chater, 2002) and the dual processing theories (e.g., Evans, 2003, 2008)[5]. The mental logic theory, the most influential rule-based

[4] Other important logic/rule-based theories include Verbal Substitution proposed by Störring (1908) and Ford (1995), Monotonicity Theory proposed by Geurts (2003), the extension of Weak Completion Semantics to syllogistic reasoning (Dietz, 2015; Stenning & Lambalgen, 2008), and recently a mental logic approach by Zhai (2015); Zhai, Szymanik, and Titov (2015). All these theories assume that the cognitive processes are based on the application of classical logical rules. However, only the last two theories make predictions for errors in human reasoning (though all of them proposed some explanations to erroneous responses but these proposals are somehow difficult to be verified). The Weak Completion Semantics can model some non-monotonic reasoning tasks (e.g. counterfactual conditional reasoning) and some effect in reasoning (e.g. belief-bias effect). PSYCOP (mental logic theory) and the last two proposals are implemented as computational models of syllogistic reasoning.

[5] Khemlani and Johnson-Laird (2012) classified the theories of syllogistic reasoning into three categories, namely, formal rule-based, models/diagrams/sets-based and heuristic-based. Instead of under the rule-based categories, the authors have put the Mental Model Theories under the category of models/diagrams/set-based. As this thesis focus on deductive reasoning instead of syllogistic reasoning (which is a major kind of deductive reasoning) alone, we adopt a different classification of the theories. Major theories of deductive reasoning (focusing on syllogistic and conditional reasoning) would be presented in this chapter (and later chapters).

theory, suggests that reasoners use some kind of inherent mental deduction rules to obtain a conclusion; while the mental model theory suggests that reasoners form mental models of the premises for drawing the conclusion (Byrne & Johnson-Laird, 2009; Johnson-Laird & Byrne, 1991). These theories propose that people reason rationally, basically, and wrong application of rules, inadequate or inaction in the search for counterexample, or the limitation of cognitive resources causes failure to produce logical responses. The probabilistic approach is based on re-analyses of empirical human reasoning data to explain participants' rationality or irrationality (which is a kind of ad-hoc proposal). The dual-process theories of reasoning suggests that we have two different types of processing, one solves problems relying on "intuition" and the other solves problems based on more controlled analytical means (loading on the working memory resources). Moreover, there is a recent view that there is no universal processing mechanism, instead the reasoners choose different strategies according to the task demand, their own capacities and abilities, experience and training (Roberts & DelVecchio, 2000). They may even employ some non-logical heuristic strategies for some more demanding reasoning tasks (e.g., Gilhooly, Logie, & Wynn, 2002b).

There are mainly four different views on the difficulty of reasoning problems. According to the heuristic-based theories, the difficulty mainly depends on the surface features of the problems. While the rule-based theories suggest that the difficulty depends on the number of steps/rules and the difficulty of the rule itself (as in mental logic theory) or the number of models (as in mental model theory) required to solve the problem, respectively. The third proposal is the interpretation theories which suggests that the difficulty depends on how people represent the problem, therefore individual differences is an important factor in reasoning performances. The fourth view, conclusion identification theories, proposes that reasoners represent the problem accurately but problem occurs during the conclusion construction process.

1.2 Mental Logic Theory

The mental logic theory suggests that humans translate the premises into logical form. Reasoning is through the manipulation of mental representations, which resemble the sentences in natural language, by applying syntactic rules of inference (rule of logic; first-order formal inference rules) (e.g., Rips, 1994). Together with some Gricean implicature rules to get rid of the existence problem of the relevant entities in logic, Rips has implemented the theory in a computational model called PSYCOP. It includes working on some formal representation with abstract logic rules. It is also stated that the implicature rules are not transitive to avoid some contradictory predictions. To prevent huge searching space which can be caused by reasoning with merely the forward rules (to derive conclusions from premises), some backward rules (to generate sub-goals) were then included. As it is based on formal logic rules, the conclusions will be always valid and no erroneous performance will be predicted. Different “versions” of this theory differ in the exact rules involve in the reasoning process. The theory explains the differences in the difficulty of problems in terms of the number of rules has to apply or the difficulty in applying particular rules. The necessary rules involved in solving the problem confounds the number of different steps required. The type of inference rules and their implicatures determine the type of error. Rips proposed that erroneous responses generated by human reasoners are due to some problems in the recognition, retrieval, or/and application of the formal rules. As the whole reasoning process is supposed to be syntactic in nature, no effect of the content (semantics) on reasoning is predicted.

However, the theory still cannot adequately explain or prove some human reasoning behaviors, especially the invalid ones. For example, for conditional reasoning, only MP and MT can be proved by formal rules. MP can be drawn directly with the formal rule of inference but the proof of MT requires several more steps, using reductio ad absurdum[6]. AC and DA inferences cannot be proved by mental logic, but they are sometimes drawn by humans.

1.3 Mental Model Theory

On the contrary, the mental model theory (MMT) proposed by Johnson-Laird and collaborators hypothesizes that reasoning involves the semantics of the arguments. The representation of the problem involves iconic spatial "diagrams", namely the mental models. There are three stages in the reasoning process: premise interpretation, premise integration and conclusion validation. In the first stage, both the terms and their categorical relations are converted to abstract tokens (abstract concepts like negation or quantifiers are represented as well). Reasoners then form provisional integrated initial mental model(s) of the syllogism in the second stage. Mental models are assumed to be constructed and manipulated in working memory (WM). Some people may construct an image of what a model represents from a certain point of view, more specifically in the form of spatial arrangements (Knauff & Johnson-Laird, 2002), which implies that mental models are spatial[7] (and iconic) in nature and draw on the resources of

[6] It means finding a contradiction to the supposition of the antecedent, p. Reasoners firstly suppose p after reading the two premises and then find that the conclusion, q, (by applying the MP inference rule on the supposition) and not-q (the second/minor premise) are incompatible and thus reject the supposition of p using reductio ad absurdum and finally conclude not-p.

[7] See the visual impedance effect studies (e.g. Castañeda & Knauff, 2013; Knauff & Johnson-Laird, 2002; Knauff & May, 2006) for details. The two-streams hypothesis -- the ventral and dorsal streams for visual and spatial locational processing respectively (Goodale & Milner, 1992; Smith et al., 1995), and the presence of two separate subsystems

the visuospatial sketchpad (VSSP) of WM particularly. It is also assumed that a major component of reasoning is nonverbal. In order not to overload the WM, reasoners try to present as little information as possible in the mental models and thus only what is true according to the premises is represented, but not what is false [8](e.g., Johnson-Laird & Byrne, 2002; Johnson-Laird, Girotto, & Legrenzi, 2004), and some linguistic properties of the premises may not be represented. The third stage is the only real deduction process. It involves a model-based search for counterexamples to the initial model. Error in performance is mainly due to insufficient search for the counterexamples or mistakes in the construction of the models. Most of the syllogisms are multiple-model problems, solving them requires the construction of more than one mental model. For example, the AA1 syllogism is a single-model problem:

Premise 1: All As are Bs

Premise 2: All Bs are Cs

Mental model for AA1:

```
a  b  c
a  b  c
   b  c
      c
…
```

The valid conclusion “All As are Cs” can be derived from the above model. AE2 is an example of three-model syllogisms:

Premise 1: All Bs are As

---

of the visuospatial sketchpad may suggest two possible representation formats of the reasoning problems during processing, namely spatial or visual. Knauff and Johnson-Laird (2002) found evidence supporting the spatial nature of the mental representation of reasoning problems.

[8] Minimum information is represented explicitly due the limitation of cognitive resources, but some other information is represented implicitly in “mental footnote” (in Johnson-Laird’s term).

Premise 2: No Cs are Bs

Mental model for AE2:

<u>Model 1</u>	<u>Model 2</u>	<u>Model 3</u>
b a	b a	b a
b a	b a	b a
a	a c	a c
c	c	a c
c		

The valid conclusion “Some As are not Cs” can only be drawn when the three models are constructed. The other invalid conclusions will be drawn if the initial model is not fleshed out, for example “No As are Cs” according to Model 1. While for conditional reasoning problems, there are three metal models for the full explicit models.

The problem is more difficult if more models have to be constructed as reasoners have to maintain and consider all of them to justify the validity of the conclusion. This process loads on cognitive resources and takes time (Bell & Johnson-Laird, 1998; Copeland & Radvansky, 2004) and thus leads to errors and inefficiency in reasoning when reasoners have limited WM capacity to represent and manipulate mental models (e.g., Johnson-Laird & Byrne, 1991) and/or limited time to search for counterexamples. MMT explains deviation of human reasoners from the normative logically correct performance by inaction or failure in the search of counterexamples and fleshing out of the initial mental model.

For example, in conditional reasoning, after reading the conditional statement “If p then q”, the initial model is formed as:

p   q

…

Then, a putative conclusion is formed with what is true according to the model. In this case, when p is given as the fact, q is derived as the conclusion (modus ponens inference). In cases where for example not-q is given, the model has to be fleshed out. This process is carried out when it is necessary and cognitive resources allow. Through the process of searching for counterexamples, the fully fleshed out model is as:

p   q

not-p   q

not-p   not-q

Khemlani and Johnson-Laird have recently developed a computational implementation of the mental model theory, known as mReasoner, which includes also the plausibility of the use of heuristic in reasoning.

1.4 Probabilistic Heuristic Model and Dependence and Independence Model

The Probability Heuristic Model (PHM) for syllogisms (Chater & Oaksford, 1999; Oaksford & Chater, 2007) is one of the most well-developed heuristic-based theories[9] of human

[9] Atmosphere and Matching hypothesis are the other two well-known reasoning theories based on heuristics. They will be introduced with more details in later chapters. However, they explain syllogistic reasoning performance only. The essence of these theories mainly hypothesizes that the selection or endorsement of the conclusion quantifier according to the two premise quantifiers. The hypotheses differ in their preferred orders of the conclusion quantifier when the two premise quantifiers are not identical. Only the PHM provides prediction regarding the conclusion

reasoning. It is one of the theories that build on the idea of Bayesian modeling. The idea was based on the observation that empirical reasoning performance usually deviates from normative responses. Oaksford and Chater think that models or theories based on logic approach cannot explain the human data adequately as human reasoning performance is not essentially logical. They then proposed a model based on probability calculus and generated five ordered heuristic rules to apply in reasoning (three for conclusion generation and two for evaluating the tentative conclusion). A good fit of the prediction from the model and human data was found in the meta-analysis (for syllogistic reasoning).

Different from the two theories above that the cognitive mechanisms for deductive reasoning is general for all kinds of reasoning, the probabilistic heuristic model proposes that different specific reasoning mechanisms are used for different kinds of reasoning. One main core rationale of this theory is based on the uncertainty of the world, that any general factual claim is "defeasible", for example, the atypical exemplars such as dolphins are not fish but mammals – warm-blooded marine animals. Oaksford and Chater argued that models to explain how people reason should firstly explain how people make inferences in daily life and then explain participant's performance in laboratory setting.

---

direction (i.e., a-c or c-a). The atmosphere hypothesis suggests that reasoners seem to select the conclusions with the "mood" which agrees with the "atmosphere" of the two premise quantifiers while the matching hypothesis suggests that reasoners choose the more conservative quantifier (the one representing fewer entities) among the two premise quantifiers. The order of conservativeness of the four quantifiers is as: E >O = I »A. Both theories do not predict any NVC responses. The atmosphere hypothesis gives predictions of reasoner's responses but does not explain why. It does not explain the difficulty neither. While the matching hypothesis (Wetherick & Gilhooly, 1990) suggests that reasoning errors are due to illicit "conversion" such as converting "All A are B" to "All B are A".

The theory is developed based on probability/uncertainty in information generalization process and involves the rational analysis by Anderson (1990) which is about information gain and belief updating: the more the information available, the lower the uncertainty. In other words, more improbable statement is more informative. They applied this principle to syllogistic reasoning by calculating the "informativesness" of statements quantified by the four quantifiers, A, I, E, and O (based on conditional probability and rarity: All (> Most > Few) > Some> No >> Some…not). The probabilities of each of the 4 statements is true are calculated, neglecting the content of concrete problems. They proposed the following five ordered heuristic rules for syllogistic reasoning, namely G1: min-heuristic, G2: p-entailments, G3: attachment-heuristic, T1: max-heuristic, and T2: o-heuristic. O-conclusions are the least frequent ones (c.p. in MMT, all O-conclusion syllogisms are difficult problems because they require the construction of 2 or 3 mental models). PHM observes and can predict the differences within 2 and 3-model problems that MMT doesn't. However, e.g., Roberts and Sykes (2005) found little evidence for the rule T2.

Oaksford et al. (2000) also proposed the Dependence Model and Independence Model for conditional reasoning akin to the PHM for syllogistic reasoning. It is also a Bayesian model about how people interpret a conditional and reason about it. Instead of interpreting "p implies q" in the classical logical sense (material implication), conditional reasoning can be modeled as the conditional probability of q given p, i.e., P(q|p) for MP, not-p given not-q for MT, p given q for AC and not-q given not-p for DA. The probability of the conclusion is calculated according to the probability of the entities in some equations. A specific conditional probability is accepted only if the computed value is above a given threshold. The best fitting parameter values for the equations were then found by substituting them with the values of 0.1, 0.3, 0.5, 0.7 and 0.9 respectively.

1.5 Dual-Process Theories

The dual process[10] theories focus on a different aspect of the cognitive mechanism involved in reasoning. The core of the theories concerns about two kinds of processing systems/processes in (general) cognitive processes, rather than the actual steps or rules in reasoning. Since the article of Wason and Evans (1975), an increasing number of authors have proposed that there are 2 types of processes (processing systems) when people reason (e.g., Evans, 2003, 2006; Evans & Over, 1996; Evans & Stanovich, 2013a, 2013b; García-Madruga, 1983, 1989; Gigerenzer & Goldstein, 1996; Sloman, 1996; Stanovich, 2004; Stanovich & West, 2000). Type 1 processes (also known as System 1) refers to the unconscious, associative, intuitive and rapid processes which give outputs that may be prone to the bias of common sense, beliefs and previous experience. It is relatively undemanding of cognitive resources and independent of fluid intelligence. Responses from Type 1 processes are quick but errors are, sometimes, inevitable. The matching heuristic process is one of the Type 1 processes. Type 2 processes (also known as System 2) are thought to be conscious, analytical, rule-based, slow and more cognitive resources demanding. They operate with effort and control and develop over time in humans.

To solve complicated problems successfully, reasoners have to go beyond the superficial Type 1 output, discard it and engage in Type 2 processing[11] (through cognitive decoupling and mental stimulation). In possible conflict resolution, several hypotheses of a control system or

---

[10] See also the heuristic-analytic theory of reasoning in Evans (2006); and heuristic and analytic processes in Schroyens, Schaeken, Fias, and d'Ydewalle (2000).

[11] We adopt the default-interventionist structure (Evans, 2007) in this thesis, though there are other proposals of how the two systems/types of processes work together, such as the parallel-competitive architecture (Sloman, 1996; Smith & DeCoster, 2000).

mechanism regarding the shift from Type 1 to Type 2 processing have been proposed (Evans, 2009; Stanovich, West, & Toplak, 2011). They involve a tripartite structure, a System 3, which deactivates Type 1 processing (System 1). However, the mechanism of this shift is still under debate. Alter, Oppenheimer, Epley, and Eyre (2007) proposed that System 2 processes are activated by metacognitive experiences which is evoked when reasoners have to solve problems of greater difficulty or disfluency, such as problems printed in a difficult-to-read font (see also Thompson et al., 2013). Besides, Thompson (2009) proposed the “feeling of rightness” hypothesis for the shift of the type of processes as well. In everyday life, people tend to accept the output of Type 1 processing and only activate Type 2 processing in some special situations, such as being explicitly instructed to reason logically (Evans, 2006; Verschueren, Schaeken, & d'Ydewalle, 2005). Due to the limitations of cognitive resources and other factors, Type 2 processing sometimes still gives wrong responses (Evans & Stanovich, 2013b).

Traditionally, Type 1 processes are thought to be context-based, while Type 2 processes are abstract and context-free. One illustration of the interaction between Type 1 and Type 2 processing in everyday reasoning can be observed in the belief bias effect. This happens when common belief and logic are in conflict. For example, for syllogisms with concrete terms, in addition to the validity of the syllogism, both premises and the conclusion can agree or contradict with common beliefs. Employing automatic heuristic processes for reasoning is more effective by demanding less cognitive resources (Evans, 2003; Sloman, 1996). It is thus usual that people prefer to use common sense heuristics over logical processes. As a result, they tend to commit the mistake of accepting invalid believable conclusions but rejecting valid unbelievable ones.

Evans and Over (2004) proposed the Suppositional Theory to explain conditional reasoning. It is a hybrid theory with the application of probability assumption akin to the independence theory and dual process theory, together with some rules in the field of pragmatics. Like the independence and dependence models, it emphasizes the cases where the probability of the consequent is high given the probability of the antecedent for MP; and similarly for the other three inferences. Contextual effect found in a vast amount of studies in conditional reasoning can be explained by pragmatic inferences. Finally, the theory has a dual system incorporated: while immediate inferences (by System 1) are solely drawn by the probability account, System 2 inferences are possible and lead to deductively valid answers.

References

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569.

Anderson, N. (1990). Spatial pattern of recent sediment and diatom accumulation in a small, monomictic, eutrophic lake. *Journal of Paleolimnology, 3*(2), 143-160.

Bell, V. A., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science, 22*(1), 25-51. doi:10.1016/S0364-0213(99)80034-2

Braine, M. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review, 85*(1), 1.

Braine, M., & O'Brien, D. P. (1998). *Mental logic*: Psychology Press.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science, 23*(3), 247-303. doi:10.1016/S0364-0213(99)00008-7

Byrne, R. M. J., & Johnson-Laird, P. N. (2009). `If'and the problems of conditional reasoning. *Trends in Cognitive Sciences, 13*(7), 282-287.

Castañeda, L. E. G., & Knauff, M. (2013). *Individual Differences, Imagery and the Visual Impedance Effect.* Paper presented at the CogSci.

Chater, N., & Oaksford, M. (1999). The Probability Heuristics Model of Syllogistic Reasoning. *Cognitive Psychology, 38*(2), 191-258. doi:10.1006/cogp.1998.0696

Copeland, D., & Radvansky, G. (2004). Working memory and syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 57*(8), 1437-1457. doi:10.1080/02724980343000846

Dietz, E.-A. (2015). A computational logic approach to syllogisms in human reasoning. *Bridging* *the Gap between Human and Automated Reasoning*, 17-31.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10).

Evans, J. S. B. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*(3), 378-395.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol., 59*, 255-278.

Evans, J. S. B. (2009). How many dual-process theories do we need? One, two, or many?

Evans, J. S. B., & Over, D. E. (1996). Rationality in the selection task: Epistemic utility versus uncertainty reduction.

Evans, J. S. B., & Over, D. E. (2004). *If*: Oxford: Oxford University Press.

Evans, J. S. B., & Stanovich, K. E. (2013a). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science, 8*(3), 223-241.

Evans, J. S. B., & Stanovich, K. E. (2013b). Theory and metatheory in the study of dual processing: Reply to comments. *Perspectives on psychological science, 8*(3), 263-271.

Ford, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition, 54*(1), 1-71. doi:10.1016/0010-0277(94)00625-U

García-Madruga, J. A. (1983). Un modelo general sobre el razonamiento silogístico: Doble procesamiento y fase de comprobación con verificación. *Revista de Psicología General y Aplicada*.

García-Madruga, J. A. (1989). Comprensión e inferencia en el razonamiento silogístico. *Cognitiva, 2*(3), 323-350.

Geurts, B. (2003). Reasoning with quantifiers. *Cognition, 86*(3), 223-251. doi:10.1016/S0010-

0277(02)00180-4

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review, 103*(4), 650.

Gilhooly, K. J., Logie, R., & Wynn, V. (2002b). Syllogistic reasoning tasks and working memory: Evidence from sequential presentation of premises. *Current Psychology, 21*(2), 111-120. doi:10.1007/s12144-002-1007-x

Johnson-Laird, P. N. (1989). Mental models. In M. I. Posner (Ed.), *The Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology, 50*(1), 109-135.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*: Lawrence Erlbaum Associates, Inc.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review, 109*(4), 646.

Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111*(3), 640.

Knauff, M., & Johnson-Laird, P. N. (2002). Visual imagery can impede reasoning. *Memory & Cognition, 30*(3), 363-371.

Knauff, M., & May, E. (2006). Mental imagery, reasoning, and blindness. *The Quarterly Journal of Experimental Psychology, 59*(1), 161-177.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*(4), 608.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*: Oxford University Press.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(4), 883.

Oaksford, M., Roberts, L., & Chater, N. (2002). Relative informativeness of quantifiers used in syllogistic reasoning. *Memory & Cognition, 30*(1), 138-149. doi:10.3758/BF03195273

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*: MIT Press.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological Bulletin, 126*(1), 3.

Roberts, M. J., & Sykes, E. D. (2005). Categorical reasoning from multiple diagrams. *The Quarterly Journal of Experimental Psychology Section A, 58*(2), 333-376.

Schroyens, W., Schaeken, W., Fias, W., & d'Ydewalle, G. e., ry. (2000). Heuristic and analytic processes in propositional reasoning with negatives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1713-1734.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and social psychology review, 4*(2), 108-131.

Stanovich, K. E. (2004). Metarepresentation and the great cognitive divide: A commentary on Henriques"“Psychology Defined”. *Journal of clinical psychology, 60*(12), 1263-1266.

Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and Brain Sciences, 23*(05), 701-717.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). The complexity of developmental predictions from dual process models. *Developmental Review, 31*(2), 103-118.

Stenning, K., & Lambalgen, M. (2008). *Human reasoning and cognitive science*. Cambridge, MA: MIT Press.

Störring, G. (1908). *Experimentelle untersuchungen über einfache Schlussprozesse*: W. Engelmann.

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. *two minds: dual processes and beyond. Oxford University Press, Oxford*.

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128*(2), 237-251.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. e., ry. (2005). A dual-process specification of causal conditional reasoning. *Thinking \& Reasoning, 11*(3), 239-278.

Wason, P. C., & Evans, J. S. B. (1975). Dual processes in reasoning? *Cognition, 3*(2), 141-154.

Zhai, F. (2015). *Toward Probabilistic Natural Logic for Syllogistic Reasoning.* (Unpublished master thesis). Universiteit van Amsterdam, Amsterdam, Netherlands.

Zhai, F., Szymanik, J., & Titov, I. (2015). *Toward a probabilistic mental logic for the syllogistic fragment of natural language.* Paper presented at the Proceedings of the 20th Amsterdam Colloquium.

Chapter 2

Individual Differences in Reasoning:

Role of Working Memory and Development

2.1 Working memory and reasoning

Reasoning is one of the complex cognitive processes which require several more fundamental processes. As mentioned in the first chapter, both the mental model theory and dual-process theories emphasize the importance of cognitive resources, particularly working memory (e.g., Barrouillet & Lecas, 1999; García-Madruga, Gutiérrez, Carriedo, Luzón, & Vila, 2007) in reasoning[12] explicitly.

The multicomponent model of working memory proposed by Baddeley and colleagues is the most influential one. They proposed a model with four components, namely the central executive, the phonological loop, the visuospatial sketchpad and the episodic buffer (Baddeley, 2000; Baddeley & Hitch, 1974). The central executive acts as the managing unit which supervises the integration of information and coordinates the two “slave systems” (the phonological loop and visuospatial sketchpad, which act as short-term storage units for information) and other cognitive processes. The phonological loop stores phonological information while the visuospatial sketchpad stores visual and spatial information, which can be further divided into the visual subsystem, for visual information such as color and shape, and the spatial subsystem, for information related to location or space (Sima, Schultheis, & Barkowsky,

[12] There were also proposals of individual differences in reasoning strategy that some people reason with a verbal and propositional while others reason with visuospatial representation (e.g. Ford, 1995; Bacon et al., 2007).

2013).

According to the mental model theory, both temporary storage of the mental model(s) and reasoning processes require the working memory. Johnson-Laird suggested that the limitation of working memory was the major cause of erroneous reasoning output. While for dual-process theories, Evans and Stanovich (2013a); (Evans & Stanovich, 2013b) suggested that the degree of involvement of working memory (WM) in the reasoning process is one of the main distinctions between Type 1 and 2 processes. Reasoners with lower working memory capacities are more likely to employ the automatic Type 1 processes to solve the problems. Reasoners with high working memory capacities (and other executive function abilities) may or may not employ Type 2 processes according to the circumstance (such as thinking style of the participant and the task demand). Individual differences in working memory capacities have been shown to be associated with syllogistic reasoning performance (e.g. Gilhooly, Logie, Wetherick, & Wynn, 1993, (Markovits, Doyon, & Simoneau, 2002); as well as syllogistic problems with belief-bias content), relational reasoning (e.g., Castañeda & Knauff, 2013), and conditional reasoning (Barrouillet & Lecas, 1999; Markovits, Doyon, & Simoneau, 2002; Toms, Morris, & Ward, 1993). For example, studies employing a dual-task paradigm have consistently reported a role for the central executive and verbal WM in syllogistic reasoning (Bacon, Handley, & Newstead, 2003; Gilhooly, Logie, & Wynn, 2002a; Gilhooly, Logie, & Wynn, 1999). While the mental model theory emphasizes the importance of spatial ability (also as suggested by the visual impedance effect), limited number of studies have found positive correlation of visuospatial working memory and reasoning performance (e.g. Markovits, Doyon, & Simoneau, 2002). It seems that among the two different working memory capacities, the verbal one is more important in reasoning. This may due to the fact that a large part of the reasoning process involves semantics, especially the “pre-logical” processes, that problems must be comprehended and transformed to certain representation before the real reasoning processes. Given the fact that reasoning problems are usually quite complex in terms of the linguistic form, usually having at least two premises (and sometimes an additional conclusion), quite some information has to be stored in the verbal working memory during reasoning. The importance of verbal working memory is thus inevitable. In this thesis, besides the effect of verbal working memory capacity (which was tested in all the four studies), we put special emphasis on the spatial working memory capacity that we tested two different, one simple and one complex, visuospatial working memory tasks in a study with adults, a complex task for the next study in adults, and a simple task for the second study in preadolescents (as the complex visuospatial task may be too difficult for the participants). Besides, we tested the importance of the central executive capacity as well, and as it is also reflected in the complex task of both visuospatial and verbal working memory. Please see later chapters for the details.

2.2 Development of reasoning in childhood

For reasoning, both the increase in working memory and metacognitive abilities facilitate the reasoning ability of children and preadolescents. It was found that working memory capacities of children from 4 to 15-year-old increase linearly in each of its components (Ambridge & Wearing, 2004). While the metacognitive ability is a late-developing skill, usually emerging around preadolescence (Moshman, 2004; Pillow, 2002), children before this age are expected to lack the full control of the shift from superficial reasoning to analytic reasoning (e.g., Santamaría, Tse, Moreno-Ríos, & García-Madruga, 2013). Several metacognitive factors may affect the changes in motivation which shift from the automatic system 1 processes to the cognitive resources loading system 2 processes in reasoning.

Besides, reasoners always are asked to assume that the premises are always true (‘‘it holds everywhere and never changes’’) in deductive reasoning, this concept is developed from early childhood (about 7 years old) but still subject to later progress (Gauffroy & Barrouillet, 2011; Markovits, 2014; Miller, Custer, & Nassau, 2000). In some difficult cases, they must inhibit their background knowledge or belief, for example “All squares are circles”. Drawing the logically valid conclusion in reasoning problems with concrete content sometimes requires the inhibition of knowledge or belief, especially in the belief-bias problems (which will be explained in later chapters). Developmental studies, such as of conditional reasoning, support that reasoning in children and preadolescents is affected by the background knowledge and belief and the pattern is rather complex. It is difficult for very young children to inhibit their background knowledge in reasoning (Markovits & Thompson, 2008). While inhibiting the content of the problems is necessary for drawing the valid conclusion in conflict problems (in which the validity and the belief of the problem is in conflict with each other), it is easier for children to draw a valid conclusion for problems with concrete premises than abstract (until early adulthood (Markovits & Lortie-Forgues, 2011)) ones (with causal ones in between, which cannot be observed consistently until around 10 to 12 year-old (Janveau-Brennan & Markovits, 1999)). The progress of the development of reasoning with concrete premises until middle adolescence is related to working memory, retrieval efficiency, and inhibitory capacity (e.g. Markovits & Barrouillet, 2002; Simoneau & Markovits, 2003).

Only until preadolescence (around 11-12 years), they begin to understand the concepts of necessity and validity of logical conclusions and can reason logically (Markovits, 2014; Moshman, 1990). For conditional reasoning, the endorsements of the invalid AC and DA inference decrease with age (while adults still commit these mistakes). The development progress in three interpretational stages: conjunctive, (defective) biconditional and finally conditional (e.g. Gauffroy & Barrouillet, 2011). A conjunctive interpretation means that the child considers that only the p q cases make the conditional true and the other three cases making conditional false as according to the mental model theory, young children can only construct the initial model, p q, out of the fully explicit model of three mental models. With the development, adolescents begin to be able to flesh out the initial model and add the additional model not-p not-q (biconditional: p q and not-p not-q). The stage when they consider the p q cases as making the conditional true, not-p not-q cases as indeterminate, but the two other cases remaining false is called defective biconditional. Only until late adolescence or adulthood can the reasoners reach the conditional stage when the full three models of a conditional can be constructed.

Compared to conditional reasoning, there are few developmental studies on syllogisms. This may due to the complexity (most of the syllogisms are different problems – of multiple mental models) and quantity (64 problems in total) of syllogistic problems. Bara, Bucciarelli, and Johnson-Laird (1995) found reliable improvement of syllogistic performances with age for children of 9-10 years old, adolescents and adults. The 9 to 10-year-old participants could give valid conclusions to one-model syllogisms well above chance only. They found also the importance of working memory in the reasoning performance in a regression analysis.

Reference

Bacon, A. M., Handley, S., & Newstead, S. E. (2003). Individual differences in strategies for syllogistic reasoning. *Thinking & Reasoning, 9*(2), 133-168. doi:10.1080/13546780343000196

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation, 8,* 47-89.

Bara, B. G., Bucciarelli, M., & Johnson-Laird, P. N. (1995). Development of syllogistic reasoning. *American Journal of Psychology, 108*(2), 157.

Barrouillet, P., & Lecas, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning, 5*(4), 289-302.

Castañeda, L. E. G., & Knauff, M. (2013). *Individual Differences, Imagery and the Visual Impedance Effect.* Paper presented at the CogSci.

Evans, J. S. B., & Stanovich, K. E. (2013a). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science, 8*(3), 223-241.

Evans, J. S. B., & Stanovich, K. E. (2013b). Theory and metatheory in the study of dual processing: Reply to comments. *Perspectives on psychological science, 8*(3), 263-271.

García-Madruga, J. A., Gutiérrez, F., Carriedo, N., Luzón, J. M., & Vila, J. O. (2007). Mental models in propositional reasoning and working memory's central executive. *Thinking & Reasoning, 13*(4), 370-393.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. Developmental Psychology, 40(2), 177.

Gauffroy, C., & Barrouillet, P. (2011). The primacy of thinking about possibilities in the development of reasoning. *Developmental Psychology, 47*(4), 1000.

Gilhooly, K., Logie, R., & Wynn, V. (2002a). Syllogistic reasoning tasks and working memory: Evidence from sequential presentation of premises. *Current Psychology, 21*(2), 111.

Gilhooly, K. J., Logie, R., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & Cognition*, *21*(1), 115-124.

Gilhooly, K. J., Logie, R., & Wynn, V. (1999). Syllogistic reasoning tasks, working memory, and skill. *European Journal of Cognitive Psychology, 11*(4), 473-498.

Janveau-Brennan, G., & Markovits, H. (1999). The development of reasoning with causal conditionals. *Developmental Psychology, 35*(4), 904.

Markovits, H. (2014). On the road toward formal reasoning: Reasoning with factual causal and contrary-to-fact causal premises during early adolescence. *Journal of Experimental Child Psychology, 128*, 37-51.

Markovits, H., & Barrouillet, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review, 22*(1), 5-36.

Markovits, H., Doyon, C., & Simoneau, M. (2002). Individual differences in working memory and conditional reasoning with concrete and abstract content. *Thinking & Reasoning, 8*(2), 97-107.

Markovits, H., & Lortie-Forgues, H. (2011). Conditional reasoning with false premises facilitates the transition between familiar and abstract reasoning. *Child development*, *82*(2), 646-660.

Markovits, H., & Thompson, V. (2008). Different developmental patterns of simple deductive and probabilistic inferential reasoning. *Memory & Cognition, 36*(6), 1066-1078.

Miller, S. A., Custer, W. L., & Nassau, G. (2000). Children's understanding of the necessity of logically necessary truths. *Cognitive Development, 15*(3), 383-403.

Moshman, D. (1990). Rationality as a goal of education. *Educational Psychology Review, 2*(4), 335-364.

Moshman, D. (2004). From inference to reasoning: The construction of rationality. *Thinking & Reasoning*, *10*(2), 221-239.

Pillow, B. H. (2002). Children's and adults' evaluation of the certainty of deductive inferences, inductive inferences, and guesses. *Child development*, *73*(3), 779-792.

Santamaría, C., Tse, P. P., Moreno-Ríos, S., & García-Madruga, J. A. (2013). Deductive reasoning and metalogical knowledge in preadolescence: a mental model appraisal. *Journal of Cognitive Psychology, 25*(2), 192-200.

Sima, J. F., Schultheis, H., & Barkowsky, T. (2013). Differences between spatial and visual mental representations. *Frontiers in psychology, 4*(240). doi:http://doi.org/10.3389/fpsyg.2013.00240

Simoneau, M., & Markovits, H. (2003). Reasoning with premises that are not empirically true: evidence for the role of inhibition and retrieval. *Developmental Psychology, 39*(6), 964.

Toms, M., Morris, N., & Ward, D. (1993). Working memory and conditional reasoning. *The Quarterly Journal of Experimental Psychology, 46*(4), 679-699.

Chapter 3

Objectives, Methodology and Results

The main objective of this series of studies was to examine individual difference in deductive reasoning as a function of working memory, cognitive abilities and development of them. Four studies were performed with both adults and preadolescents.

3.1 Objectives

Studies with Preadolescents: Study 1 & Study 2

The main objective of the two studies was to explore how the working memory and other cognitive abilities/capacities and their developments affect the reasoning performances. For study 1, we focused on (verbal) working memory capacity and metadeductive knowledge which is the main basis for the employment of Type 2 processes. Preadolescence is a developmental period in which children are completing the acquisition of both faculties. While for study 2, we extended the cognitive ability and capacity measures to fluid intelligence (as measured by k-bit), inhibitory control (Stroop task) and comprehension ability, and we included also a measure for visuospatial working memory (Corsi) and the central executive (n-Back) as potential predictors of reasoning performances of different reasoning tasks. We investigated if there is any developmental or/and interaction effect of these measures as well.

Studies with Adults: Study 3 and Study 4

We focused on working memory and inhibitory control for the two studies with adults as the development of metacognitive capacity is expected to be completed at this age. As our participants were all university students, we do not expect a big difference in their fluid intelligence and thus we did not include the kbit task. Only syllogistic tasks were performed in the two studies to narrow down the research space. In study 3, we explored the importance of working memory in generation and evaluation tasks of syllogistic reasoning. Morley, Evans, and Handley (2004) suggested that the reasoning process in conclusion generation tasks is more likely to be forward reasoning (i.e. from the premises to the conclusion) while that in conclusion verification tasks is more likely to be backward reasoning (i.e. conclusion driven). Also, the provision of the conclusion helps in the construction of the mental model(s) of the premises (Evans, Handley, & Harper, 2001; Hardman & Payne, 1995; Klauer, Musch, & Naumer, 2000). Research on syllogistic tasks found different results from these two tasks, for example, figural effect of the premises and a stronger belief bias effect was found in the generation tasks than in verification tasks (c.f. Stupple & Ball, 2005, 2007). Besides, Hardman and Payne (1995) argued that verification tasks of three-model problems with valid conclusions can be regarded as single or two-model problems while those with invalid single-model problem might be regarded as multiple-model due to the additional need of finding a counterexample. We aimed to examine the above claims, specifically if WM was a significant factor in generation task but not verification task in study 3. Besides, we aimed to address especially the claim of the importance of visuospatial working memory according to mental model theory (as unlike verbal working memory, fewer studies have found the importance of visuospatial working memory capacities in reasoning). We included a simple and a complex WM measure for visuospatial working memory, a complex task for verbal working memory and a task for the central executive. The simple task is expected to rely more on the slave system of the visuospatial memory only while the two complex tasks load on the central executive in addition to the slave systems. In study 4, besides the WM, we examined also the importance of inhibitory control on syllogistic reasoning. The use of heuristics and whether the semantic details of the syllogism are processed or represented were also investigated. We used conflict problems that the inhibition of the employment of heuristic is necessary for drawing the valid conclusion to examine the importance of inhibitory control in this study.

3.2 Methodology

Study 1 and Study 2

We tested a large population of preadolescents in the seventh and eighth grades with metadeductive and deductive reasoning tasks for both propositional and syllogistic reasoning (4 tasks in total) and a verbal working memory task for Study 1. The metadeductive task for propositional reasoning examines the evaluation of propositional attitudes, i.e. whether someone was telling the truth or not. While the one for syllogistic reasoning asked the participants to judge whether the conclusion was necessarily true, possible, or impossible. For the propositional deductive reasoning task, two conditional problems (Modus Tollens and Denying of Antecedent) and two inclusive disjunction problems (one affirmative and one negative) were used. Three classical syllogistic conclusion generation problems were used. Participants were asked to write down a valid conclusion. The three problems included one single-mode, one two-model and one three-model problems. The problems for all the tasks were selected so that they covered a wide range of difficulty. Whereas some of them could be solved by Type 1 processes and others require Type 2 processes.

We tested the same four reasoning tasks and included more measures for the fluid intelligence, inhibitory control, comprehension skills, and visuospatial and central executive of working memory.

Study 3

In this study, we tested both the generation and verification (also known as evaluation) task of syllogistic reasoning. For the generation task, participants were asked to write down their own conclusion; while they were asked to evaluate if the given conclusion was valid or not (yes/no) for the verification task. As mentioned before, we included four tasks for different working memory spans.

Study 4

Only evaluation task of syllogistic reasoning was employed in this study. We included two kinds of conflict problems and no-conflict problems. Participants were asked to perform a lexical decision task right after each syllogistic problem to examine whether there was any inhibition on the terms in the conclusion of the syllogism. In addition, they had to complete three complex working memory tasks and the Stroop task (as a measure of inhibitory control capacity).

3.3 Results

In the first study with preadolescents, we found reliable effect of verbal working memory in reasoning performances (propositional and syllogistic reasoning). However, in the second study with more cognitive measures, fluid intelligence was the most important predictor of reasoning performances for the first grader but the working memory capacities (visuospatial, verbal and central executive) for the second graders. Regression analysis showed that the fluid intelligence was the most important predictor of the reasoning performances (except syllogistic conclusion generation task) for the first grader but the WM capacities (measured by the Reading Span task, Corsi task and nBack task) for the second graders. This supported the idea that the development of WM in preadolescence helps in the development of reasoning abilities. The results pointed to theories which suggest the involvement of WM in reasoning, like the mental model theory and dual process theory.

For the studies with adults, in the third study, WM capacity is important for generation task but not evaluation task as we hypothesized. The inhibitory control ability (as measured by Stroop task) was an important factor for solving conflict syllogistic problems. However, significant correlation with WM measures was found for no-conflict problems but not for conflict problems. It might be that conflict problems were too difficult (all are 3-model problems) and solving them required WM capacities beyond the limitations of some participants. The results suggest a dual-process explanation of syllogistic reasoning. The effect of number of models of the syllogism in the generation task and the use of atmosphere/matching heuristics, particularly in the evaluation task, were also confirmed. These results are consistent with the proposal of mental model theory.

Reference

Evans, J. S. B., Handley, S. J., & Harper, C. N. J. (2001). Necessity, possibility and belief: A study of syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 54*(3), 935-958. doi:10.1080/713755983

Hardman, D. K., & Payne, S. J. (1995). Problem difficulty and response format in syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 48*(4), 945-975.

Klauer, K. C., Musch, J., & Naumer, B. (2000). On Belief Bias in Syllogistic Reasoning. *Psychological Review, 107*(4), 852-884. doi:10.1037/0033-295X.107.4.852

Morley, N. J., Evans, J. S. B., & Handley, S. J. (2004). Belief bias and figural bias in syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 57*(4), 666-692. doi:10.1080/02724980343000440

Stupple, E. J., & Ball, L. J. (2005). *An inspection-time analysis of figural effects and processing direction in syllogistic reasoning.* Paper presented at the Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society.

Stupple, E. J., & Ball, L. J. (2007). Figural effects in a syllogistic evaluation paradigm: An inspection-time analysis. *Experimental Psychology, 54*(2), 120-127.

Experimental Series

Chapter 4

Deductive reasoning and metalogical knowledge in preadolescents: a mental model appraisal

Reference: Santamaría, C., Tse, P.P., Moreno-Ríos, S., & García-Madruga, J. (2013). Deductive reasoning and metalogical knowledge in preadolescence: A mental model appraisal. *Journal of Cognitive Psychology, 25*(2), 192-200. DOI: 10.1080/20445911.2012.743988

Abstract: In this paper we analysed the metalogical and deductive inferential ability of a wide sample (1118 participants) of seventh and eighth grade school students (12-13 years old). We used two metareasoning tasks: an evaluation of propositional attitudes and a modal syllogistic task. Two additional deductive (propositional and syllogistic) reasoning tasks were used: a propositional inference task and a syllogistic construction task. We also tested the participants' working memory spans with the Reading Span test (Daneman & Carpenter, 1980). We found a reliable effect of working memory for all the tasks, but an effect of school grade only for the metareasoning tasks. The results support the idea that metareasoning competencies make unusual progress during preadolescence. This development is crucial for individuals to engage in analytical reasoning.

Keywords: Mental models; Metalogic; Preadolescence; Reasoning.

4.1 Introduction

A symptom of cognitive development is that people exhibit a tendency to increment the number of elements they are able to keep in mind. For example, older children can maintain in their memory more words (Gathercole & Pickering, 2000), numbers (Isaacs & Vargha-Khadem, 1989), or visual patterns of higher complexity (see Gathercole, 1999 for a review; Wilson, Scott, & Power, 1987), and, with increasing age, represent concepts of higher relational complexity (Case, 1985; Halford, Wilson, & Phillips, 1998). Gathercole, Pickering, Ambridge, and Wearing (2004) studied children aged from 4 to 15 years using a battery of working memory tasks. They demonstrated that the capacity of working memory in each of its components increases linearly throughout this period. Also, reasoning psychologists have observed that when younger children are asked to list what is possible given a conditional sentence like If it rains then Sally uses an umbrella, they list only one of the true possibilities: that in which both the If-clause and the Then-clause are true (the situation in which it rains and Sally uses an umbrella). Older children add the situation in which both clauses are false (the situation in which it is not raining and Sally is not using an umbrella), and from adolescence onwards people tend to list all three true possibilities (including the situation in which the If-clause is false and the Then-clause is true: it is not raining and Sally uses an umbrella; (Barrouillet, 2011; Barrouillet, Grosset, & Lecas, 2000; Barrouillet & Lecas, 1999).

Being able to hold several possibilities in mind does not mean taking all of them into account all the time. For instance, in most cases, when understanding conditionals, adults focus on the case that even younger children can list: the case in which both clauses are true (Johnson-Laird, 2006; Johnson-Laird & Byrne, 2002). The exact procedure by which people go beyond this initial model is not known. However, we might assume that people have both the motivation and the capacity to engage in the effortful processing needed to go beyond the initial models. In developmental terms, capacity shift may be mainly determined by the enhancement of working memory span (WMS), and changes in motivation may be determined by several metacognitive factors that support the shift from a superficial reasoning based on a limited effort to analytic reasoning.

There is considerable agreement that people reason in two ways. One is rapid, automatic, and only its end product is accessible to consciousness (System 1 processes). The other way of reasoning (System 2) is slow and costly, and makes extensive use of the central working memory system (Evans, 2003, 2008; Evans & Over, 1996; Kahneman & Frederick, 2002; Sloman, 1996; Stanovich & West, 2000). Alter et al. (2007) proposed that System 2 processes are activated by metacognitive experiences. These authors manipulated their tasks so as to steer their participants into experiencing greater difficulty or disfluency, and this manipulation led them to engage in analytical processing. For example, in their fourth experiment they found that participants responded better to syllogisms when they were printed in a difficult-to-read font. It seems that experiencing difficulty is a cue that leads people to activate System 2 processes.

Fletcher and Carruthers (2012) have recently proposed three main features of System 2 which are particularly relevant to metacognition in reasoning:

(1) It is subject to intentional control.

(2) It can be guided by normative beliefs about proper reasoning methods.

(3) It overrides the unreflective responses endorsed automatically by System 1.

All these features share evaluation of one’s own reasoning as a prerequisite, and this capacity is a late-developing skill, usually emerging around preadolescence (Moshman, 2004; Pillow, 2002).

We might predict that before this age children will lack full control of their shift from

System 1 to System 2 reasoning. For this purpose, we need to make use of tasks that involve unequivocally metalogical processes. Most reasoning tasks often involve metalogical reasoning but some of them allow reasoners to solve them without the use of this skill. From the point of view of mental model theory, this posture has been defended by authors like Schroyens (2010) in his distinction between thinking from conditionals and thinking about conditionals.

The evaluation of the truth or falsity of metalinguistic predicates requires the testing of an assertion against the world or against other assertions, so that it is not possible to solve them without the participation of complex metalogical skills. In this paper, we use two kinds of genuine metalogical tasks: propositional attitudes and modal reasoning. The understanding and use of propositional attitudes (expressions of the mental posture of a person towards a proposition, e.g., John thinks that there is a circle) and modal operators (such as those of necessity, possibility, and impossibility, e.g., It is possible that there is a circle) in reasoning is a sign of metacognitive competence. Indeed, it has been argued that the relationship between them is quite close (for instance, through a dedicated mind-reading system; Carruthers, 2009). Developmentally, the concept that necessary truths ‘‘hold everywhere and never change’’ can be found in a simple way from early childhood (7 years old) but is still subject to later progress (Miller et al., 2000).

Our purpose in the present research was to explore how the two conditions of System 2 reasoning (working memory capacity and metadeductive knowledge and motivation) interact in preadolescence, a developmental period in which children are completing the acquisition of both faculties. For this purpose, we tested a large population of preadolescents in two school-age groups (seventh and eighth grades) with metadeductive and deductive reasoning (propositional and syllogistic) tasks and a WMS task. One of the metadeductive tasks entailed the evaluation of propositional attitudes (whether or not someone was telling the truth regarding an assertion, in other words whether or not the assertion is consistent with the facts). It included five items of conditional and disjunctive reasoning. For example:

Andres said he would go to typing-class or

English-class, but not both.

Finally he didn’t go to typing-class and he

didn’t go to English-class.

Did Andres tell the truth?

Another metadeductive task used three modal syllogisms where the participants were asked to evaluate each conclusion as necessarily true, possible, or impossible. For example:

All artists are gardeners

Raquel is an artist

Conclusion: Raquel is a gardener

Necessarily True/Possible/Impossible

The two additional deductive reasoning tasks were a four-item propositional inference task that included two conditional problems (Modus Tollens and Denying of Antecedent) and two inclusive disjunction problems (one affirmative and one negative); and a three-item syllogistic task in which participants were asked to generate a valid conclusion for three syllogisms of differing difficulty. The problems were selected so that they covered a wide range of difficulty. Whereas some of them could be solved by confirming the initial (System 1) conclusion, others, being difficult multiple model problems, needed an active search for counterexamples (see Johnson-Laird & Byrne, 1991 for an analysis of these tasks in terms of mental models).

In order to solve propositional and syllogistic reasoning tasks, people need to apply System 2 processes. The normative knowledge and beliefs about proper reasoning methods, in particular reasoners' knowledge about the consistency of and necessity for deductive conclusions, as shown in metadeductive tasks, are prerequisites for the activation and correct functioning of System 2 processes.

Moshman (1990) affirms that during preadolescence, around 11-12 years, people reach the third stage of development, which means that they begin to understand the concepts of necessity and validity of logical conclusions, and they can reason about the logical form of propositions from a formal logical system. Velasco and García-Madruga (1997) investigated the development of metalogical understanding and logical reasoning using categorical syllogism with preadolescents (seventh and eighth graders) and adolescents (10th and 12^{th} graders). They confirmed that a third of seventh graders (12.5 years old) were still unable to understand correctly the logical concepts that characterize the third stage of logical development. There was a gradual acquisition of metalogical understanding during preadolescent and adolescent years. Eighth graders and adolescents showed the acquisition of implicit metalogical understanding of the logical system that characterizes the third stage. According to Moshman, preadolescents and adolescents cannot think about the logical system as a whole and take it as an object of knowledge. This explicit metalogic capacity is only possible in adulthood when people reach the fourth stage of development. Thinking about the logical system allows people to grasp the relationship between natural languages and the logical system, which will be needed particularly to perform the most difficult metadeductive tasks.

In their study, Gauffroy and Barrouillet (2011) tested children and adolescents of different ages and school levels (fifth, sixth, and ninth graders) and university students. They found that the course of development for the two tasks (reasoning about possibilities and reasoning about truth values) is the same as that predicted by the mental model theory: The developmental pattern is predicted by the increasing number of possibilities or mental models that participants are able to consider in order to solve conditional problems: one model (conjunction), two models (biconditional), and three models (conditional). All this evidence leads us to expect that our metacognitive tasks will be more sensitive to school grade progress at this preadolescent age.

4.2 Method

Participants

A total of 1118 children (527 girls) from 28 different secondary schools in four provinces of Spain participated in the experiment as part of a broad Research Project on the relationship between cognitive abilities and education. All were native speakers of Spanish, and they were tested in that language. Students with any diagnostic problems were eliminated from the analysis, leaving 982 students. Of these, 541 were seventh graders in secondary school (18 ESO; mean age: 12.23, SD: 0.73) and the rest (441) eighth graders (28 ESO; mean age: 13.28, SD: 0.68).

Materials and procedure

The participants were tested individually in a quiet classroom at the same school. The problems were given in a booklet for each participant, no time limit was set, and the participants were instructed not to go back when solving the problems. The instructions encouraged the students to read the problems carefully for comprehension.

They completed four reasoning tasks and a WMS task. WMS was measured with Daneman and Carpenter’s Reading Span Test (RST; Daneman & Carpenter, 1980); Spanish version for children by Orjales, García-Madruga, and Elosúa (2010). In this task, participants are asked to read out loud a series of sentences on a computer screen and then to recall the last word of each sentence in the correct order. The sentences are very simple and easy to read, using familiar words. The task includes different levels in which the number of sentences increases progressively from two to six. There are three series of sentences at each level. Within each of these three series, a participant's performance can be (1) correct (accurate words, correct order), (2) half correct (accurate words, incorrect order), and (3) incorrect. The scoring procedure marks the level (from 2 to 6) at which participants are able to remember the words with minimum consistent performance; that is, when a participant performs at least half of the maximum: either three series of words half correct; or one series of words correct, one half correct, and one incorrect. Each performance better than the minimum consistent performance at the same or higher levels is scored by the addition of decimals. On the same level, each supplementary correct response would add two decimal points and each supplementary half correct response one decimal point. On a higher level, a supplementary correct response would add five decimal points and a supplementary half correct response, four decimal points.

The reasoning tasks included two propositional reasoning tasks and two syllogistic reasoning tasks:

Task 1: Propositional inference task. This task was framed in a cooking environment so that the premises took the form of statements in a cookery course. The instructions stressed that the conclusion selected should be true whenever the premises were true.

This task included two conditional problems* Modus Tollens (MT) and Denying of Antecedent (DA), and two inclusive disjunction problems: affirmative and negative disjunctions. The problems read as follows (*correct response):

(1) (DA) For the first dish, the teacher says:

“If I use mayonnaise, then I use lettuce.”

And then asserts:

“I do not use mayonnaise.”

What can you conclude about the other ingredient?

. They use lettuce.

. They do not use lettuce.

. Cannot form a conclusion.*

(2) (MT) For the second dish, the teacher says:

“If I use garlic, then I use oil.”

And then asserts:

“I do not use oil.”

What can you conclude about the other ingredient?

. They use garlic.

. They do not use garlic.*

. Cannot form a conclusion.

(3) (Affirmative disjunction) For the first dessert, the teacher says:

“I use honey or cinnamon, or both.”

And then asserts:

“I use honey.”

What can you conclude about the other ingredient?

. They use cinnamon.

. They do not use cinnamon.

. Cannot form a conclusion.*

(4) (Negative disjunction) For the second dessert, the teacher says:

‘‘I use chocolate or wine, or both.’’

And then asserts:

‘‘I do not use wine.’’

What can you conclude about the other ingredient?

. They use chocolate.*

. They do not use chocolate.

. Cannot form a conclusion.

Task 2: Evaluation of propositional attitudes.

This task made reference to five fictional students who said what extracurricular activities they would engage in and what they finally did. The participant’s task was to judge whether or not the character was telling the truth. Before the experimental problems, the participants were given an example with a single negated proposition. The problems read as follows (*correct response):

(1) Sonia said she would go swimming or to soccer, but not both.

Finally she did not go swimming, but she went to soccer.

Did Sonia tell the truth? Yes*/No

(2) Andres said he would go to typing-class or English-class, but not both.

Finally he didn’t go to typing-class and he didn’t go to English-class.

Did Andres tell the truth? Yes/No*

(3) Marta said that if she went to basketball then she would not go to tennis.

Finally she went to basketball and tennis.

Did Marta tell the truth? Yes/No*

(4) Paco said that if he went to chess then he would go to squash.

Finally he did not go to chess and he did not go to squash.

Did Paco tell the truth? Yes*/No

(5) Laura said that if she went to karate then she would go to the gym.

Finally she didn't go to karate but she went to the gym.

Did Laura tell the truth? Yes*/No

Task 3: Syllogistic construction task. In this task, the participants were asked to write their own conclusions for three pairs of syllogistic premises of increasing difficulty. The content referred to the occupations of people in a small village. The instructions explained that the conclusions should connect the two terms that were not repeated in the premises. A problem with no valid conclusion was used as an example. The three problems were:

(1) All skiers are vegetarians.

Some plumbers are skiers.

Conclusion?

(2) Some mechanics are teachers.

All motorists are mechanics.

Conclusion?

(3) No football player is a programmer.

All football players are athletes.

Conclusion?

Task 4: Modal syllogistic task. In this task, the participants were asked to evaluate if the proposed conclusion was necessarily true, possibly true, or impossible. The three problems were:

(1) All engineers are mathematicians.

Sergio is an engineer.

Conclusion: Sergio is not a mathematician.

Necessarily True/Possible/Impossible*

(2) All artists are gardeners.

Raquel is an artist.

Conclusion: Raquel is a gardener.

Necessarily True*/Possible/Impossible

(3) All doctors are physicists.

Luis is a physicist.

Conclusion: Luis is a doctor.

Necessarily True/Possible*/Impossible

The reasoning tasks were presented to the participants in this fixed order (Task 1, Task 2, Task 3, Task 4) after the WMS task.

4.3 Results

Grade effects on the reasoning tasks

The percentages of accuracy for each of the reasoning problems in the two grades are shown in Table 1. The overall correct response rates reached 48.85% and 52.34% for Grade 7 and Grade 8 students, respectively. A one-way ANOVA for accuracy in each of the four reasoning tasks showed no reliable differences in the propositional inference task, $F(1, 977) = 2.81$, $\eta^2 = .003$; $p = .094$, or in the syllogistic construction task, $F(1, 955) = 2.34$, $\eta^2 = .002$; $p = .127$. However, as predicted, the two metacognitive tasks were more sensitive to school grade: for the evaluation of propositional attitudes task, $F(1, 970) = 15.21$, $\eta^2 = .015$; $p = .0001$; and for the modal syllogistic task, $F(1, 975) = 7.31$, $\eta^2 = .007$; $p = .007$.

Table 1

The Percentages of Accuracy for Each of the Problems in the Two Grades

Problems and Statement		grade 1 (N=530-540)	grade 2 (N=436-441)
Propositional Inference Task			
Conditional Inferences	DA	8.3% (0.28)	10.0% (0.30)
	MT	58.5% (0.49)	56.6% (0.50)
Inclusive Disjunctions	Aff.	23.1% (0.42)	30.3% (0.46)
	Neg.	79.3% (0.41)	80.9% (0.39)
Overall Task		42.3 (0.40)	44.5 (0.41)
Eval. of propositional attitudes task			
Truth Table Excl. Disj.	"no-p q"	90.3% (0.30)	94.8% (0.22)
	"no-p no-q"	90.5% (0.29)	93.2% (0.25)
Conditional	"p q"	92.0% (0.27)	96.8% (0.18)
	"no-p no-q"	43.5% (0.50)	54.3% (0.50)
	"no-p q"	17.0% (0.38)	13.0% (0.34)
Overall Task		66.7 (0.35)	70.4 (0.30)
Syllogistic construction task			
Syllogisms	Easy	38.5% (0.49)	44.6% (0.50)
	Difficult	28.2% (0.45)	27.5% (0.45)
	Very Difficult	0.0% (0.00)	0.2% (0.05)
Overall Task		22.2(0.31)	24.1(0.33)
Modal syllogistic task			
Nec. & Possib.	Impossible	66.5% (0.47)	73.2% (0.44)
	Necessary	66.6% (0.47)	68.3% (0.47)
	Possible	38.2% (0.49)	45.6% (0.50)
Overall Task		57.1 (0.48)	62.4 (0.47)
Overall correct response rates		48.85% (0.13)	52.34% (0.12)

Note: standard deviations are shown in parentheses.

Working memory effects on the reasoning tasks

In order to test the effect of WMS on the reasoning tasks, we selected the 30% of participants with the highest WMS and the 30% with the lowest WMS in each grade according to their scores in the RST (as measured with Daneman & Carpenter, 1980: the remaining 40% of the data were discarded from the analysis). The data (presented in Table 2) were submitted to a 2 (reading span: high working memory or low working memory) x 2 (grade: Grade 7 or Grade 8) ANOVA for each of the four reasoning tasks. We found reliable effects of WMS for the four reasoning tasks. For the propositional inference task, $F(1, 454) = 15.774$, $\eta^2 = .034$; $p < .001$; evaluation of propositional attitudes, $F(1, 454) = 25.642$, $\eta^2 = .054$; $p < .001$; syllogistic construction task, $F(1, 454) = 12.814$, $\eta^2 = .028$; $p < .001$; and modal syllogistic task, $F(1, 454) = 32.998$, $\eta^2 = .068$; $p < .001$. Although the effect of school grade was tested in the previous analysis, we repeated the analysis with the subsample of participants who completed the RST. With this method, we confirmed that this subsample had the same behavior as the total sample regarding school grade. Indeed, we found the same results as with the total sample, namely, a significant effect only for the metacognitive tasks: for evaluation of propositional attitudes, $F(1, 454) = 14.688$, $\eta^2 = .032$; $p < .001$; for modal syllogistic task, $F(1, 454) = 7.604$, $\eta^2 = .017$. There was no significant interaction effect for any of the four tasks.

4.4 Discussions

Working memory span is essential for any effortful processing. The central executive is a limited capacity workspace used for both the storage and the processing of information (Baddeley & Hitch, 1974). The immediate consequence is that tasks demanding greater effort to process information leave less space for the storage of that same information, and vice versa. The effects are difficulty and error in complex tasks. For example, Daneman and Merikle (1996) found that measure that exploit the combined processing and storage capacity of working

Table 2

Reasoning Accuracy (Percentage of Correct Responses) of ohe 30% Highest and Lowest Reading Span Participants of Each Reasoning Category in the Two Grades

	grade 7		grade 8	
	High (N=115)	Low (N=115)	High (N=112)	Low (N=112)
Propositional Inference Task	46%(0.21)	39%(0.22)	49%(0.20)	41%(0.18)
Eval. of propositional attitudes task	67%(0.15)	61%(0.19)	74%(0.11)	65%(0.16)
Syllogistic construction task	23%(0.20)	17%(0.19)	27%(0.20)	19%(0.20)
Modal syllogistic task	62%(0.27)	44%(0.31)	67%(0.28)	54% (0.29)

Note: standard deviations are shown in parentheses.

memory (like the RST) are the best predictors of comprehension.

Any standard reasoning memory (like the RST) are the best predictors of comprehension. Any standard reasoning task should be considered a complex task in that sense. Consequently, it is not surprising that we found effects of memory span on all tasks. Children with greater WM capacity are less prone to error in reasoning. This result confirms previous findings (see Capon, Handley, & Dennis, 2003; De Neys, Schaeken, & d'Ydewalle, 2005a, 2005b; García-Madruga et al., 2007). Indeed, Kyllonen and Christal (1990) found correlations of between .80 and .90 between WM and reasoning ability factors (see also, for structural equation modelling studies of that relationship;Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). Also, some common neuronal processes have been identified for WM and deductive reasoning (Ruff, Knauff, Fangmeier, & Spreer, 2003).

During preadolescence, at the beginning of secondary school, relevant changes in reasoning abilities have been noted. Inhelder and Piaget (1955) claimed that between 11-12 and 14-15 years, children's concrete thinking becomes abstract and formal. These new abilities have been interpreted as due to a change in metacognitive knowledge and abilities: an advance in the capacity to differentiate logical necessity from empirical truth (Moshman, 1990).

Working memory certainly increases with age (Barrouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009; Gathercole et al., 2004), particularly during the preadolescent years. The changes detected in the present study could be due to the increase of WM, but also to formal training at school and its effect in relevant fields such as reading comprehension and mathematics, as well as the acquisition of general knowledge. It would not be too surprising if formal education had a special effect on metareasoning tasks. As Barrouillet (2011, p. 170) has recently affirmed, some developmental findings suggest that, beyond the increase in working memory capacity (Barrouillet & Lecas, 1999), the development of conditional reasoning could be based on metacognitive development.

The WMS affected all our tasks, whereas the educational grade (seventh vs. eighth) yielded a significant effect only for the metadeductive reasoning tasks, and no reliable interaction was found. It seems that the age and school grade of our participants (12-13 years) is crucial in the development of the metacognitive processes that lead to analytic reasoning and that are central in modal reasoning and the evaluation of propositional attitudes. These two tasks should be considered a measure of the participants' metadeductive capacities as they entail dawning awareness of their own and others' mental functions (Flavell, 1976, 1979; Kuhn, 2000) and the crucial role of consistency and necessity on deductive inferences.

Our results support the idea that full control of the metacognitive processes involved in reasoning is a late advance in human development. As Gauffroy and Barrouillet (2011, p. 1008) maintained, the mental model theory proposal is consistent with Piaget (1987) ideas about possibility and necessity. Piaget claimed that formal operational thinking depends on the capacity to understand and yield possibilities, and hence the capacity to consider possibilities precedes the operations on which rational thinking is based. Fully functioning metalogic is therefore a late achievement.

Also, it should be noted that in our experiment a clear rise in correct responses in the metadeductive tasks was obtained with a tiny school grade and age difference between the two groups. This might reflect the fact that our participants were at an age of change for this capacity. Improvement in other capacities such as WMS would have been too small to produce reliable differences in our experiment. These improved metareasoning aptitudes might lead them to recognize the need to flesh out their models. Therefore, the two traditional requisites for reasoners to go beyond their initial models and even search exhaustively for counterexamples -- capacity and motivation -- might need to be complemented by metareasoning knowledge. It seems that preadolescents have acquired enough WM capacity but they are still developing additional knowledge and skills. These skills are essential for full control of the situations in which the reasoner must shift from System 1 to System 2.

Awareness of the attitudes of others (as in our propositional attitudes task) together with other control and monitoring procedures would be useful adaptations at the age when members of our species are getting ready to leave the safety of their families. Consequently, it might be argued that a distinctive adaptation for metareasoning occurs during preadolescence. However, as Fletcher and Carruthers (2012) have noted, metareasoning is subject to a significant variation.

Obviously, people in different cultures and individuals with different levels of instruction in formal disciplines would have quite different metareasoning aptitudes. Consequently, for these authors ‘‘there is not a distinct and distinctive metacognitive system for monitoring and controlling one’s reasoning’’ (p. 1375). It is obvious that there must be many different ways to manage controlled reasoning. In fact the psychology of reasoning has never proposed a unitary method for experts in logic and the layperson. The fact that System 2 entails conscious reasoning makes it particularly sensitive to cultural and educational intervention, but there is still room for some motivational and attitudinal changes that make people shift from an automatic and comfortable way of thinking to an effortful and costly but, sometimes, more accurate way. According to our results in this paper, these changes are striking in preadolescence and to some extent independent of processing capacity workspace.

Reference

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569.

Baddeley, A., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation, 8*, 47-89.

Barrouillet, P. (2011). Dual-process theories of reasoning: The test of development. *Developmental Review, 31*(2), 151-179.

Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: a time-based resource-sharing model account. *Developmental Psychology, 45*(2), 477.

Barrouillet, P., Grosset, N., & Lecas, J. F. (2000). Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition, 75*(3), 237-266.

Barrouillet, P., & Lecas, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning, 5*(4), 289-302.

Capon, A., Handley, S., & Dennis, I. (2003). Working memory and reasoning: An individual differences perspective. *Thinking & Reasoning, 9*(3), 203-244.

Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences, 32*(02), 121-138.

Case, R. (1985). Intellectual development: Orlando, FL: Academic Press.

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*(2), 163-183.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422-433.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005a). Working memory and counterexample retrieval for causal conditionals. *Thinking & Reasoning, 11*(2), 123-150.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005b). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning, 11*(4), 349-381.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10).

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol., 59*, 255-278.

Evans, J. S. B., & Over, D. E. (1996). Rationality in the selection task: Epistemic utility versus uncertainty reduction.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. *The nature of intelligence,* 231-235.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist, 34*(10), 906.

Fletcher, L., & Carruthers, P. (2012). Metacognition and reasoning. *Phil. Trans. R. Soc. B,* *367*(1594), 1366-1378.

García-Madruga, J. A., Gutiérrez, F., Carriedo, N., Luzón, J. M., & Vila, J. O. (2007). Mental models in propositional reasoning and working memory's central executive. *Thinking & Reasoning, 13*(4), 370-393.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3*(11), 410-419.

Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology, 70*(2), 177-194.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*(2), 177.

Gauffroy, C., & Barrouillet, P. (2011). The primacy of thinking about possibilities in the development of reasoning. *Developmental Psychology, 47*(4), 1000.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*(06), 803-831.

Inhelder, B., & Piaget, J. (1955). *De la logique de l'enfant à la logique de l'adolescent*. Paris: Presses Univ. de France. (2e éd. 1970.).

Isaacs, E. B., & Vargha-Khadem, F. (1989). Differential course of development of spatial and verbal memory span: A normative study. *British Journal of Developmental Psychology, 7*(4), 377-380.

Johnson-Laird, P. N. (2006). *How we reason*. Oxford: University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*: Lawrence Erlbaum Associates, Inc.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review, 109*(4), 646.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment, 49*.

Kuhn, D. (2000). Metacognitive development. *Current directions in psychological science, 9*(5), 178-181.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*(4), 389-433.

Miller, S. A., Custer, W. L., & Nassau, G. (2000). Children's understanding of the necessity of logically necessary truths. *Cognitive Development, 15*(3), 383-403.

Moshman, D. (1990). Rationality as a goal of education. *Educational Psychology Review, 2*(4), 335-364.

Moshman, D. (2004). From inference to reasoning: The construction of rationality. *Thinking & Reasoning, 10*(2), 221-239.

Orjales, I., García-Madruga, J. A., & Elosúa, M. (2010). Prueba de amplitud lectora para primaria [Reading span test for primary school]. *Unpublished manuscript*.

Piaget, J. (1987). *Possibility and necessity* (Vol. 1): U of Minnesota Press.

Pillow, B. H. (2002). Children's and adults' evaluation of the certainty of deductive inferences, inductive inferences, and guesses. *Child development, 73*(3), 779-792.

Ruff, C. C., Knauff, M., Fangmeier, T., & Spreer, J. (2003). Reasoning and working memory: common and distinct neuronal processes. *Neuropsychologia, 41*(9), 1241-1253.

Schroyens, W. (2010). A meta-analytic review of thinking about what is true, possible, and irrelevant in reasoning from or reasoning about conditional propositions. *European*

Journal of Cognitive Psychology, 22(6), 897-921.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3.

Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and Brain Sciences*, *23*(05), 701-717.

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence, 30*(3), 261-288.

Velasco, J., & García-Madruga, J. A. (1997). La búsqueda de contraejemplos como estrategia metalógica. Un estudio sobre la influencia de su entrenamiento en el razonamiento silogístoco. *Infancia y aprendizaje*, *20*(80), 17-35.

Wilson, J., Scott, J. H., & Power, K. G. (1987). Developmental differences in the span of visual memory for pattern. *British Journal of Developmental Psychology*, *5*(3), 249-255.

Chapter 5

Predictors of Reasoning Performances in Preadolescence

Authors: Tse, P.P., García-Madruga, J., Moreno-Ríos, S., & Bajo, M.T.

Abstract: The reasoning performance of five different reasoning tasks of 1118 first and second grade secondary school students were found to correlate with some of the working memory (WM), fluid intelligence, inhibitory control ability and comprehension ability measures. Further regression analysis showed that the fluid intelligence, as measured by the kbit task, was the most important predictor of the reasoning performances (except syllogistic conclusion generation task) for the first grader but the WM capacities (measured by the Reading Span task, Corsi task and nBack task) for the second graders. This supported the idea that the development of WM in preadolescence helps in the development of reasoning abilities. The results pointed to theories which suggest the involvement of WM in reasoning development, like the mental model theory and dual process theory.

5.1 Research background

In the famous Japanese movies/television drama "Galileo" based on the novel "Detective Galileo", the main character, Manabu Yukawa, developed hives on his body when talking with irrational people. And thus, he strongly resisted communicating with children as he thought that they were unable to think logically. Being able to reason "logically" is an important achievement of cognitive development in adolescents. Children are thought to be easily cheated, naïve and irrational as they lack the ability to reason "analytically". They exhibit different reasoning processes and give different conclusion than adults. They cannot distinguish conclusions from premises or consequence from causes and factors and cannot notice the contradiction in their conclusions or reason with abstract ideas or hypothetically. The ability to reason develops with age, along with working memory capacities, inhibitory control and meta-logical ability. One of the most influential (classical) works on children reasoning development was the series of studies performed by Piaget and his colleagues. He proposed that cognitive development is through active experiences with the environment with "assimilation" and "accommodation" of new experiences into existing framework/schemas. Children must go through all the four stages in the order of sensorimotor intelligence, preoperational thinking, concrete operational thinking and finally formal operational thinking. Children reach a specific stage at similar age. He proposed that the adult-like reasoning process is achieved in adolescence. There have been many debates on his theory since 1960s, though there is a consensus that children archive adult-like reasoning ability in adolescence. In this study, we focus on the reasoning development in early adolescents or preadolescents and its factors based on two theories reformulated from the mental model theory (MMT) and the dual-process theory.

Reformulation of the mental model theory and dual-process theory

Long after the theoretical innovation of Piaget's theories on thinking and reasoning (Inhelder & Piaget, 1955, 1964; Piaget, 1953, 1970), researchers have come up to several different theories on reasoning development before adulthood. One of the most important theories is the one proposed by Markovits and Barrouillet in 2002, which is a reformulation of Johnson-Laird's mental model theory (e.g., Johnson-Laird & Byrne, 2002) on adult reasoning (e.g., Barrouillet, Markovits, & Quinn, 2002; Markovits, 2004; Markovits & Barrouillet, 2002). The theory is based on conditional reasoning studies. They have found that children's performance of the truth table task showed a developmental trend from "conjunctive" in grade six to “defective biconditional” in grade nine and then to “defective conditional” interpretation in adults (Barrouillet & Lecas, 1998, Gauffroy & Barrouillet, 2011). According to the mental model theory (e.g., Johnson-Laird & Byrne, 2002), the difficulty of the reasoning tasks depends on the number of mental model the reasoner has to construct in order to solve the problem correctly. The more mental models have to be constructed, the higher load on the working memory (WM), as cognitive resources are needed to construct and hold the mental models in the WM. This pattern agrees with the view that the increase in working memory capacities and other cognitive abilities with age helps in the construction and maintenance of the necessary mental models during reasoning. The pattern also agrees with the sequence of the mental models to be constructed: the conjunctive pattern corresponds to the single initial model, the defective biconditional pattern to the two-model respond and finally the defective conditional pattern to the three-model full model response which represent all the possibilities:

p   q

not-p not-q

not-p q

Barrouillet, Gauffroy, and Lecas (2008) found results supporting the above suggestion when they asked the participants of different grades to evaluate the truthiness (or probability) of the four possible logical cases of conditional reasoning (true-table) defective truth-table. They found that 8-year-old children could not determine the truthiness of the case when the antecedent is false (conjunctive response). Adolescents gave a biconditional response that they think only “p q” is true but both “p not-q” and “not-p q” are false, with not-p not-q intermediate (e.g., Evans & Over, 2004; Gauffroy & Barrouillet, 2009). Generally speaking, the performance of modus ponens and modus tollens, the two valid inferences, reaches near ceiling level by late adolescence.

Gauffroy and Barrouillet (2011) distinguished reasoning judgement between possibilities and about truth values. It was found that the performance for the possibilities tasks got improvements at an earlier age than the truth table task, in general with 3 years of lag for the latter task. They tested the developmental trend for conditional (conjunctive, biconditional and conditional). It was found that children aged about 12 (average age of 11.7 years, grade 6) gave biconditional responses when they were asked to reason with possibilities, but they still gave conjunctive responses to the truth-value task. Three years later (with average age of 15.5 years, third year of secondary school), the differences diminished.

The dual-process theory of cognitive development proposed by Klaczynski (2004, 2005) is another important theory. In line with Evans' dual-process model (e.g., Evans, 2006, 2007, 2010; Evans & Over, 1996; Evans & Stanovich, 2013a) which suggests two distinct types/systems of cognitive processes which develop and work together, namely the "experiential" system (System 1) and the "analytic" system (System 2), children show less "experiential" responses with age and the development of the "metacognitive operators" with age can invoke the analytic system. The analytic system can inhibit the experiential system and extract the decontextualized representations for analytical processes which can produce a logical output which is free of biases due to the content effect. Handley and his colleagues (e.g. Handley, Capon, Beveridge, Dennis, & Evans, 2004) found that the reasoning of belief-biased problems in 10 year-old children can be predicted by an inhibitory control measure (as well as WM capacity measure), suggesting that inhibition of real-world knowledge/heuristics which hinder the application of inferences rules was important for analytical processes (see also, Kuhn & Franklin, 2006).

Barrouillet (2011) has proposed a dual-process theory which integrates the dual-process theories and the mental model theory (MMT) for conditional reasoning. He proposed that the System 1 output, a default model representing the relation between the antecedent and consequent, is essentially the initial mental model of the MMT. The modus ponens response of young children is what the initial mental model represented of the conditionals. For the "defective biconditional" and "defective conditional" patterns in older children, adolescents and adults, alternative models have to be constructed and represented and this involves the decoupling processes – System 2 processes. As both the MMT and the dual-process theory postulate, the working memory capacity is crucial in this process. With the development of WM with age, children are able to construct, maintain and manipulate more and more additional mental models with age. As mentioned above, the second model to be constructed and maintained would be "not-p not-q" and then the "not-p q" as the third model, which results in the MP and AC responses in young children and then DA in older children and finally the MT response in adolescents and adults respectively.

There are fewer developmental studies on syllogistic reasoning and probabilistic reasoning, most of them found results accountable by the dual-process theories (probabilistic reasoning: e.g., Chiesi, Primi, & Morsanyi, 2011) or mental model theories (syllogistic reasoning: e.g., Johnson-Laird, Oakhill, & Bull, 1986).

Metacognition

Gauffroy and Barrouillet (2011) explained the lap in development when reasoning with truth-values instead of possibilities that in the first case a coherent understanding of the conditional is required. In order to evaluate the "truth-value" of a sentence, reasoners must consider all the possible cases and judge whether it is true or not for each of them. For the authors, the judgment of truth is a meta ability because it requires the integration of all the possibilities in one unique and coherent structure which acquired around 15 years old. This proposal seems to be consistent with Moshman's theory regarding the development of metalogical understanding. Preadolescents, after around 11 years old, can make inference thinking in possibilities. However, they will have difficulties to think about the truth values as an explicit metalogic is needed and it is reached in the fourth level of development, around 15 years old.

The development of the metalogical understanding is crucial for reasoning with System 2 intervention. It is commonly accepted that metacognitive processes are needed for the shift from the default System 1 to System 2, in specific situations (e.g., Thompson, 2009, 2010). It usually emerges around the age of 10-12 with the ability to reason logically among propositions that are hypothetical or even false (Markovits & Vachon, 1989; Moshman, 2004; Moshman & Franks, 1986) or without contextual support (Markovits, Schleifer, & Fortier, 1989). Markovits and Vachon (1989) found that belief bias effect decreases across late childhood and adolescence, which supports the progression in separation of form and content in deductive reasoning tasks. In addition, Barrouillet (2011) suggested that the development of metacognitive processes would "enrich" System 1 output because the metacognitive experiences such as "feeling of retrieval fluency" and "feeling of rightness" (FOR) (FOR; Thompson, 2010) that accompany memory retrieval would be refined, leading to "increasingly rich default models". This agrees with Piaget's view that reflexive thinking is achieved in adolescence.

Stanovich (2009) has proposed two "minds" within System 2, namely the algorithmic mind and the reflective mind. The algorithmic mind is responsible for cognitive decoupling which enable us to construct and maintain several different representations in the WM and serial associative cognition. The capacity of this mind is related to the fluid intelligence. The reflective mind is responsible of initiating the System 2 intervention and the inhibition of heuristic responses (System 1 responses). This may be the system which is responsible for the FOR phenomenon. We, therefore, include the k-bit task to measure the fluid intelligence of children.

Factors affecting reasoning performance

The effect of experiences in reasoning can be both positive and negative as with the number of heuristics acquired, children may tend to use heuristics rather than consuming their increased cognitive resources for normative responses. The availability of more cognitive resources and the development of metacognition abilities does not necessarily lend to more normative responses, if he/she does not choose to employ them. Failure to give normative responses (especially in conflict conditions) may due to either the lack of cognitive resources so that children fail to inhibit the irrelevant materials and incorrect responses; or that children do not aware of the existence of the alternatives (requiring metacognition); or that children opt for the easier processes to solve the problems. And thus, the findings of many previous studies are rather controversial. In this study, we try to find the importance of (the development of) these two factors, namely the development of cognitive resources and metacognition, in different reasoning tasks in preadolescents. Previous studies mainly focused on only one type of reasoning tasks, or only the effect of executive functions or age/grade. We tried to achieve a more integrated study of these important factors.

Many studies have pointed to the importance of WM in reasoning (e.g., Barrouillet et al., 2000; Barrouillet & Lecas, 1999; Handley et al., 2004), as mentioned above. Some belief-bias effect studies pointed to the importance of the inhibition control to successful decontextualisation of conflict problems (with the believability of the problem being incongruent with the logical validity) for correct responses, as in adult studies. Other factors include the fluid intelligence (Stanovich, 2009) and the comprehension skills.

Besides the effect of the development of executive functions and fluid intelligence, the ability to reason rationally is also consolidated through education that certain relevant logical rules have to been taught to children, especially for probabilistic reasoning. Adolescents are taught the rules and concepts about proportions and percentages, and the calculation of some probability mathematical questions in the classroom. Researchers have showed that without formal instruction, even intelligent adults cannot answer the probabilistic reasoning problems correctly because it is greatly affected by heuristics. In Spain, basic rules and concepts of these knowledge are taught to children around 12 and 13 years old (the ages of our participants). As what we found in Santamaría, et al. (2013), the education level of children is a factor affecting the reasoning performance of the two meta-cognitive tasks (truth table and modal syllogistic task) but not the propositional inference and syllogistic construction tasks. We predicted that the grade level did not affect the performance of the probabilistic reasoning task as well.

In this study, we aimed to evaluate the importance of different factors, including WM, fluid intelligence and comprehension ability, in five different reasoning tasks. All the above theories point to the importance of WM capacities in the development of reasoning abilities. We used three WM tasks, namely reading span task, Corsi visuospatial WM task and n-back task, in this study (see Miyake & Shah, 1999 for the review of working memory models). Besides, we used the kbit task, which measures the fluid intelligence to evaluate the hypothesis of Stanovich's suggestion regarding the involvement of the "algorithmic mind" and the Stroop task to evaluate the hypothesis about the importance of the inhibition capacity in different reasoning tasks.

In order to have an overall evaluation of different reasoning task in early adolescents/preadolescents, we employed five different reasoning tasks: two conditional reasoning tasks, an inference task and a metalogical task (evaluation of propositional attitudes); two syllogistic tasks, a deductive and a metalogical task (modal syllogistic task); and a probabilistic reasoning task. We included conditional inferences and disjunctions in the propositional tasks. Problems of different levels of difficulty were included in each task. The difficulty depends on the number of alternatives to be considered: single-model (easier) and multiple-model (more difficult as predicted by mental model theory) problems. We expected that the three-model problem was extremely difficult for our participants and a low percentage of correct responses for the two-model problems. We expected also a greater developmental improvement in the two metalogical tasks, due to the acquisition of metacognitive ability at this stage, than the inference tasks.

We have demonstrated a reliable effect of the verbal working memory capacity (as measured by the reading span task) in the two propositional and syllogistic reasoning tasks in Santamaría, et al. (2013). We included more WM measures in this study for a more comprehensive study of the relationship between different modules of the WM and different reasoning tasks. In addition, probablistic reasoning was also tested in this study.

Aims of the study

Preadolescence has been proposed as a critical period in the development of cognition, and particularly of reasoning. We tested students of the first and second grades of secondary schools with tasks to evaluate the relationship between their executive operations and their ability to solve different reasoning problems. Problems with different levels of difficulty (based on the predictions of the mental model theory) were tested. Though there have been many studies on the relationship between reasoning performances and executive functions, most of these studies used a composite measure for the reasoning performance and the reasoning tasks they used are general tasks, such as the I-S-T 2000R (Buehner, Krumm, & Pick, 2005) and raven task, which some researchers argued that they were similar to WM tasks in some sense (Süß et al., 2002). These tasks measure general reasoning abilities and most studies have found that WM measures are good predictors or have a high correlation with these general measures. However, we predicted that different reasoning tasks should involve different processing and thus different executive functions and WM components should have different level of involvement in different reasoning tasks. Therefore, we included three types of reasoning tasks which are massively used in adult reasoning studies, namely propositional, syllogistic and probabilistic reasoning. In additional, we have two kinds of propositional and syllogistic tasks – a metalogical and an inference task. We used the three classic WM tasks which have been shown in many different studies as good measures of WM capacities.

Our first aim is to verify the predictions of the difficulties of different problems in each task according to the mental model theory. Regarding the difficulty of the problems, we included problems of different number of mental models in each task. We expected a higher percentage of correct responses for single-model than multiple-model problems, as multiple-model problems are supposed to be more difficult. There were two two-model and two three-model problems for the propositional inference task; three single-model, one two-model and one three-model problem in the evaluation of propositional attitude task; one single-model, one two-model and one three-model problem in the syllogistic construction task; two single-model and one two-model problems in the modal syllogistic task.

For the propositional inference task, we expected a greater difficulty for the conditional inferences than the disjunctions, in preadolescence. While for evaluation of propositional attitudes, we expected a greater difficulty for the conditional with negation than conditional problem than disjunctions. The multiple-model problems were expected to be more difficult than the single-model problems in the syllogistic conclusion construction task and modal syllogistic task.

Our second aim is to check the developmental pattern of different reasoning tasks. As mentioned above, we expected a greater improvement for the two metalogical tasks than the two deductive tasks and probabilistic reasoning task with age. Though the difference is only one year, we still expected a significant improvement.

Our third aim is to evaluate the predictive power of WM capacities, fluid intelligence, inhibitory control and comprehension ability for the performance of different reasoning tasks (and check if there is any developmental effect):

Reasoning and WM measures (reading span, Corsi span and nBack task)

The development of the WM in preadolescence would improve the reasoning performance in propositional and syllogistic reasoning. We predicted a higher predictive power of the WM measures in second graders but not for the probabilistic reasoning task (though there were controversial results regarding which component of WM predicts a specific reasoning task).

Reasoning and fluid intelligence

Before the enhancement of the WM in preadolescence, we predicted the fluid intelligence as the most significant predictor of the reasoning performances.

Reasoning and Stroop task

The ability to inhibit irrelevant information and heuristic System 1 is crucial for the more difficult problems, as mentioned above. However, it would not be the most importance predictor.

The inhibitory control is expected to play a more important role for problems that reasoners have to inhibit the irrelevant information from the long-term memory/background knowledge (as in belief-bias studies). For example, reasoners have to inhibit their knowledge about some careers in the syllogistic construction task as not “all the skiers are vegetarians” in reality and “some football players may be a programmer” (see Appendix A for a full list of the problems we tested in this study). They have to think hypothetically or ignore the content for such kinds of problems.

Reasoning and comprehension measures

Comprehension ability is the foundation to understand the problems. The skill is essential for readers to connect text features and his knowledge and experiences, with several processing levels of different components. It involves reasoning and metalogical processes as well because readers have to construct an interpretation of the text with numbers of arguments based on the information by the text, reader’s knowledge and writer’s intention. Readers have to reason consistently to detect possible misunderstandings during the process. Furthermore, according to the mental model theory, construction of mental models involves semantics. Correct understanding of the problem is a prerequisite of reasoning. Thus, we expected a high correlation of the comprehension skills with the reasoning performances.

Moshman (1990) suggested that children of 11 to 12-year-old reach beginning of the stage when can reason logically from the premises disregarding the content. We tested first and second grade secondary school students as they are of the age of 11 to 13. We think that the grade level rather than the age can predict better the reasoning performance of children. Therefore, our forth aim was to test the effect of grade levels on different reasoning tasks. We divided the reasoning tasks into metalogical tasks and inference tasks to test the interaction between the development of metacognition and executive function, fluid intelligence and comprehension skills in reasoning.

5.2 Method

Participants

A total of 1,118 children (527 girls) from 28 different secondary schools in five provinces of Spain participated in the experiment as part of a broad Research Project on the relationship between cognitive abilities and education. All of them were native speakers of Spanish and they were tested in that language. Students with any diagnostic problems were eliminated from the analysis, leaving 982 students. Of these, 541 were first graders in secondary school (1º ESO; Mean age = 12.23; SD = .73) and the rest were (441) second graders (2º ESO; Mean age = 13.28; SD =. 68).

Reasoning Materials

We used "Prueba de razonamiento para secundaria**"** (Reasoning test for secondary school) which was developed by García-Madruga et al. (2009) as the test material for the reasoning tasks. The test consists of five reasoning tasks. See Appendix A for all the problems.

Task 1. Propositional inference task. This task was framed in a cooking environment so that the premises took the form of statements in a cookery course. The problems were on the ingredients used in the preparation of certain dishes and desserts on a basic cooking class. In each problem, there was a sentence that relates two ingredients followed by a statement about if one of the ingredients mentioned in the sentence was used or not. The instructions stressed that the conclusion selected should be true whenever the premises were true. This task included two conditional problems (*'if A then B')*: modus tollens (MT) and denial of antecedent (DA); and two inclusive disjunction problems ('A or B, or both'): affirmative and negative disjunctions.

The responses to the various statements were classified into three types: symmetric responses when the conclusion has the same polarity (positive or negative) as the premise, categorical propositional responses (no conclusion), and asymmetric responses when the conclusion has a different polarity as the premise.

In the ‘if then’ statements, the valid conclusion to the modus tollens (MT) inference is the symmetrical response while that in denial of antecedent (DA) is the non-propositional answer “no conclusion”. For disjunctive statements, the valid conclusion for the affirmative problem is the ‘no conclusion’ response, while that for the negative problem is the asymmetric response.

Task 2. Evaluation of propositional attitudes (Truth Table task). This task involved judging whether or not each of the five students was telling the truth about what extracurricular activities they would engage in (first premise). What they finally did were provided as the fact (second premise). Before the problems, the participants were given a single negated proposition as an example, as follows:

Sonia said that she would go swimming or play football, but not both. (First Premise, the claim)

Finally she played football. (Second premise, the fact)

Did Sonia tell the truth? YES / NO (the correct answer is “YES”)

Task 3a. Syllogistic construction task. In this task, participants were asked to provide the conclusions for three syllogistic problems of increasing difficulty. The contents were about the occupations of people in a small village. Participants were asked to write down conclusions that should connect the two terms which were not repeated in the premises. A problem with no valid conclusion was used as an example. The first problem is a single-model problem with the correct response agrees with heuristics (atmosphere or matching). The second problem is a two-model problem with “no valid conclusion” as the correct responses. While the third problem was a three-model problem with the correct response disagrees with heuristics. System 2 processes must be employed for the correct conclusion.

Task 4. Modal syllogistic task. We used three problems of similar type -- Premise 1: All A are B. Premise 2: C is A/B. Conclusion: C is/is not A/B. The terms in the problems were common occupations. In this task, participants were asked to evaluate if the given conclusion was *necessarily true, possibly true* or *impossible*. For example:

Premise 1: All engineers are mathematicians

Premise 2: Sergio is an engineer

Conclusion: Sergio is not a mathematician

Response: Necessarily True / Possible / Impossible* (Impossible is the correct response)

Task 5. Probabilistic reasoning task. This task included three problems related to probability and three possible answers. The first problem was taken from Tversky and Kahneman (1974) which require the rule that “departures from population are more likely in small samples”, a rule that children learn in statistical lessons. The required rule of the second problem is “likelihood of independent and equiprobable events” while people give the incorrect response due to gambler’s fallacy. The third problem is on base-rate fallacy that some events are more likely to happen, for example a hit positive than a false alarm positive cancer test result. See appendix A for the three problems.

Procedure

Reasoning tasks. The participants were tested individually in a quiet classroom in their own school. The problems were given in a booklet for each participant, no time limit was set and the participants were instructed not to go back after solving each problem. Students were instructed to read the problems carefully.

Other tasks. The other tasks were completed in a counterbalance order in another room within two days. The tasks included a Spanish version of Daneman and Carpenter's Reading Span Test (García-Madruga et al., 2013), Corsi Visuospatial WM Span Test (Corsi, 1972), nBack task (Kirchner, 1958), Stroop task, kBit (a standard test for fluid intelligence) and a visual text comprehension task.

Reading Span task. In this task, participants had to remember the last word of sentences presented in sets of increasing number of sentences, from sets of two sentences to six. One sentence was presented in the center of the screen each time. Participants had to read aloud each sentence and remember the last word of all the sentences in each test set. Participants were asked to recall the last words in the order according to the presentation sequence of the sentences after each set. There were three sets at each level of 2 sentences to six sentences. The level was presented in an increasing order.

Corsi visuospatial WM task. A matrix of nine squares was presented on the screen. One of the squares was colored green and was shown for 500ms for each screen. The screens with different square colored green were shown consecutively when a key was pressed. The task started with the level of two screens. Participants had to click the mouse to indicate the two previously colored squares in the order they were presented. Gradually, the program increased the difficulty by showing more screens in each set. The maximum level was nine squares. Each level was repeated for three times. The participant obtained that level if he/she got at least two out of the three trials correct.

nBack task (Kirchner, 1958). The participant saw letters presented at a fixed rate at the center of the screen. After each letter, participant had to indicate whether that letter was the same as the previous one (key L) or different (key S). For example, for the sequence SLLSS at the one-Back level, participants should press: different, different, same, different, same. After the one-Back trials, participants were asked to press key L if the letter was the same as the previous *two* letter (two-Back). Thus, for the SLSLL sequence, the correct answers are: different, different, same, same, different. The maximum level was the three-Back trials. Participants were asked to respond as fast as possible without making mistakes. For one-back, the first letter is always considered as different and so the first two letters for two-back and first three letters for three-back.

KBIT task. We used the fluid scale (nonverbal) of the Kaufman Brief Intelligence Test which is a Matrices subtest that measures nonverbal skills and problem-solving abilities.

Stroop task. Stimuli in different colors were shown at the center of the screen. The participant had to read aloud the color of the stimuli. Five different colors were used. The stimuli consisted of a row of '%%%%%%' and words which relate to color or other events. Each of the colors repeated for five congruent trials (the color agrees with the meaning of the word), five incongruent trials (the color disagrees with the meaning of the word) and five control trials (non-color words or '%%%%%%'). Participants were asked to respond as quickly as possible and avoid making any mistakes.

Visual text comprehension task. The participant had to read texts that appear at the center of the screen. Each text contained four paragraphs. When the participant finished reading a paragraph, he/she had to press the space bar to read the next paragraph and so on until the last paragraph of the text. After this, four sentences (one at a time) appeared at the center of the screen and the participant had to press the L key if they thought that the statement was true according to the text or S key if otherwise. The participant read a minimum of two texts and a maximum of four, depending on the accuracy of the previous answers.

5.3 Results

Results of the 15 reasoning problems, except the three probabilistic reasoning problems were published in Santamaría et al. (2013). In this case, the problems were classified into single-model problems versus multiple-model problems to test our hypotheses. Please find the percentages of correct responses of the 18 reasoning problems in table 3. All the single-model metalogical problems (both propositional and syllogistic) had a percentage of correct responses (much) higher than 50%. First and second graders showed similar performance pattern. For the propositional inferences task, conditional problems were more difficult than disjunctive problems. Among the two conditional problems, DA was more difficult than the MT problem. The negative disjunctive had a higher percentage of correct responses. The final conditional problem in the evaluation of propositional attitude task seemed very difficult for our participants. It was a three-model problem (if p then q, no-p, q: true) that even many adults give incorrect answer. Half of our participants were able to answer the two-model problem (if p then q, no-p no-q: true) correctly, showing that they have already reached the "defective biconditional" stage.

Regarding the difficulties, the single-model problems were significantly easier than multiple-model problems, overall: $t(981) = 56.201$, $p < .001$; evaluation of propositional attitude task: $t(981) = 54.675$, $p < .001$; modal syllogistic problems: $t(981) = 13.185$, $p < .001$.

As expected, grade two students had a higher percentage of correct responses than first graders in general, especially in the two metalogical tasks. We did not find significant differences for three of the eight problems only. For the three inference tasks (propositional, syllogistic and probabilistic), we found significant higher percentages of correct responses for the propositional affirmative disjunction and the easiest probabilistic problem, but no significant difference was found for the three syllogistic problems.

Table 3

Descriptive Statistics of Each Reasoning Problem

MM	grade 1 and 2 students (aggregated)				grade 1 students				grade 2 students			
		Mean	SD	N		Mean	SD	N		Mean	SD	N
3	Prop_DA	.09	.29	980	Prop_DA	.08	.28	540	Prop_DA	.10	.30	440
2	Prop_MT	.58	.49	980	Prop_MT	.59	.49	540	Prop_MT	.57	.50	440
3	Prop_Daff	.26	.44	979	Prop_Daff	.23	.42	540	Prop_Daff	.30*	.46	439
2	Prop_Dneg	.80	.40	980	Prop_Dneg	.79	.41	540	Prop_Dneg	.81	.39	440
1	TruT_ED1nq	.92	.27	977	TruT_ED1nq	.90	.30	538	TruT_ED1nq	.95**	.22	439
1	TruT_ED2nn	.92	.28	979	TruT_ED2nn	.91	.29	539	TruT_ED2nn	.93	.25	440
1	TruT_Con3pq	.94	.23	979	TruT_Con3pq	.92	.27	538	TruT_Con3pq	.97**	.18	441
2	TruT_Con4nn	.48	.50	978	TruT_Con4nn	.43	.50	538	TruT_Con4nn	.54**	.50	440
3	TruT_Con5nq	.15	.36	976	TruT_Con5nq	.17	.38	536	TruT_Con5nq	.13	.34	440
1	Sil_1	.41	.49	969	Sil_1	.39	.49	532	Sil_1	.45	.50	437
2	Sil_2	.28	.45	968	Sil_2	.28	.45	532	Sil_2	.28	.45	436
3	Sil_3	.00	.03	966	Sil_3	.00	.00	530	Sil_3	.00	.05	436
1	Sil_NyP_1Im	.70	.46	978	Sil_NyP_1Im	.67	.47	538	Sil_NyP_1Im	.73*	.44	440
1	Sil_NyP_2Ne	.67	.47	980	Sil_NyP_2Ne	.67	.47	539	Sil_NyP_2Ne	.68	.47	441
2	Sil_NyP_3Po	.42	.49	980	Sil_NyP_3Po	.38	.49	539	Sil_NyP_3Po	.46*	.50	441
	Prob_1	.17	.37	977	Prob_1	.17	.37	537	Prob_1	.16	.37	440
	Prob_2	.48	.50	978	Prob_2	.44	.50	537	Prob_2	.53**	.50	441
	Prob_3	.44	.50	979	Prob_3	.41	.49	538	Prob_3	.46	.50	441

Note. t-tests of the results for grade 1 and 2 students: *$p < .05$; ** $p<.01$. MM = number of mental model according to mental model theory; SD = standard deviation; N = number of participants. Prop = propositional, TruT = truth table, Sil = syllogistic, Prob = probabilistic. See Appendix A for the details of the problems. Problems are listed in the same order as Appendix A.

Table 4

Correlations of Reasoning Tasks with the WM, Stroop, kbit and Comprehension measures for Grade 1 and 2 Students (aggregated, selected items)

	ReadSpan	Corsi	nBack	z_WM	Stroop	kbit_Vocab	kbit_Mat	kbit	compre
Prop	.207**	.125**	.169**	.249**	.038	.196**	.221**	.234**	.121**
TruT	.296**	.270**	.230**	.357**	.141**	.326**	.325**	.347**	.216**
Sil_Q1	.181**	.179**	.208**	.233**	.063	.195**	.231**	.217**	.157**
Sil_NyP	.248**	.186**	.195**	.330**	.049	.248**	.330**	.310**	.147**
Prob_Q2	.091*	.070	.094*	.150**	.093**	.179**	.175**	.192**	.144**

Note. Prop = propositional inference task, TruT = evaluation of propositional attitudes, Sil_Q1 = the first syllogistic construction problem (single-model), Sil_NyP = modal syllogistic task, Prob_Q2 = the second question of the probabilistic reasoning task. z_WM = aggregated z-scores of the working memory tasks, compre = accuracy of the comprehension task.

Correlation analysis

In both school grades, the performances of all the reasoning tasks correlated significantly with most of our measures, except the probabilistic reasoning task with Corsi measure and the Stroop task with propositional inference, syllogistic inference and modal syllogistic task. For first graders, only the correlations between the probabilistic task and the WM measures; and Stroop task with any of the reasoning tasks were *not* significant. However, the pattern for the second graders was slightly different. For second grader, the Stroop task correlated significantly with the syllogistic construction task and the evaluation of propositional attitudes task. The reading span measure correlated with the probabilistic task but the comprehension ability measure did not correlate with the modal syllogistic task. Except the probabilistic task, there was a slight trend that the correlation coefficients were larger for grade 2 students than grade 1 students. For students of both grades, it seemed that the correlation between the two metalogical tasks with the cognitive measures were larger (except the modal syllogistic task with Stroop for grade 1 students; and with kbit and comprehension accuracy for grade 2 students). See table 4-6 for the details.

Table 5

Correlations of Reasoning Tasks with the WM, Stroop, kbit and comprehension measures for Grade 1 Students (selected items)

	ReadSpan	Corsi	nBack	z_WM	Stroop	kbit_Vocab	kbit_Mat	kbit	Compre
Prop	.222**	.152**	.202**	.241**	.023	.215**	.297**	.297**	.150**
TruT	.215**	.252**	.221**	.318**	.082	.313**	.350**	.384**	.192**
Sil_Q1	.121*	.186**	.194**	.174*	.027	.108*	.187**	.193**	.124**
Sil_NyP	.267**	.193**	.182**	.274**	.005	.236**	.353**	.330**	.181**
Prob_Q2	.042	.104	.088	.145*	.083	.157**	.137**	.181**	.138**

Note. Prop = propositional inference task, TruT = evaluation of propositional attitudes, Sil_Q1 = the first syllogistic construction problem (single-model), Sil_NyP = modal syllogistic task, Prob_Q2 = the second question of the probabilistic reasoning task. z_WM = aggregated z-scores of the working memory tasks, compre = accuracy of the comprehension task.

Regression analysis

We decided to exclude five very difficult problems with less than 30% correct responses (highlighted in red in Table 3) in the regression analysis as the low percentage of correct responses may ruin the statistics (participants might get the correct response with just guessing): the propositional denial of antecedent (DA) inference(9%), biconditional DA and affirmation of consequent (AC) in the evaluation of propositional attitude task (15%), the two multiple-model syllogistic problems (28%, 0%), and the first probabilistic reasoning problem (17%).

We used the stepwise method for all the regression analyses. Participants with any missing data were excluded. For grade 1 students, the KBIT measure was the most significant predictor of both the propositional inferences and modal syllogistic tasks (b = .34, $t(174)$ = 5.94, $p < .001$ and b = .40, $t(174)$ = 5.69, $p < .001$ respectively). KBIT explained a significant proportion of variance in the two tasks, R^2 change = .169, $F(1,174)$ = 35.34, $p < .001$ and R^2 change = .157, $F(1, 174)$ = 32.38, $p < .001$. However, for the evaluation of propositional attitude task (Truth Table), the WM composite was the most significant predictor (b = .24, $t(174)$ = 2.81, $p < .006$). It explained a significant proportion of variance of the task with the KBIT measure, R^2 = .407, $F(1,173)$ = 17.17, $p < .001$. Please refer to table 7 for the details.

Table 6

Correlations of Reasoning Tasks with the WM, Stroop, kbit and comprehension measures for Grade 2 Students (selected items)

	ReadSpan	Corsi	nBack	z_WM	Stroop	kbit_Vocab	kbit_Mat	kbit	compre
Prop	.175**	.084	.124*	.235**	.048	.172**	.131**	.170**	.078
TruT	.340**	.310**	.215**	.383**	.213**	.322**	.278**	.292**	.214**
Sil_Q1	.197**	.169*	.222**	.270**	.101*	.288**	.277**	.233**	.181**
Sil_NyP	.209**	.174*	.208**	.374**	.100	.253**	.295**	.284**	.092
Prob_Q2	.109*	.017	.085	.111	.094	.187**	.208**	.191**	.128**

Note. Prop = propositional inference task, TruT = evaluation of propositional attitudes, Sil_Q1 = the first syllogistic construction problem (single-model), Sil_NyP = modal syllogistic task, Prob_Q2 = the second question of the probabilistic reasoning task. z_WM = aggregated z-scores of the working memory tasks, compre = accuracy of the comprehension task.

Table 7

Regression Analysis for the Reasoning Performances (selected items):

Stepwise with z_WM, KBIT, Stroop and Overall Accuracy of the Comprehension Task as Potential Predictors for Grade 1 Students

Reasoning task category		adj.R^2	R^2 change	predictor	β	t-vale	*p*=
Propositional Reasoning							
	Inferences	.164	.169	kbit	.411	5.944	<.001
	Truth Table	.156	.133	z_WM	.239	2.810	.006
			.032	kbit	.219	2.581	.011
Syllogistic Reasoning							
	Syllogisms	.027	.032	z_WM	.180	2.408	.017
	modal	.152	.157	kbit	.396	5.691	<.001
Probabilistic reasoning		.033	.038	z_WM	.195	2.629	.009

For grade 2 students, the WM composite (the aggregated results of the working memory tasks) was the most significant predictor of all the tasks, except probabilistic reasoning (b = .22, $t(131) = 2.52$, $p = .013$, $b = .19$, $t(131) = 2.03$, $p = .045$, $b = .25$, $t(131) = 2.94$, $p = .004$ and b = .35, $t(131) = 4.25$, $p < .001$ respectively). The WM composite explained a significant proportion of variance in the modal syllogistic task, R^2 change = .115, $F(1, 131) = 18.10$, $p < .001$. More importantly, the WM composite together with the Stroop inhibition index were good predictors

Table 8

Regression Analysis for the Reasoning Performances (Selected Items):

Stepwise with z_WM, KBIT, Stroop and Overall Accuracy of the Comprehension Task as

Potential Predictors for Grade 2 Students

Reasoning task category		adj.R^2	R^2 change	predictor	β	t-vale	*p*=
Propositional Reasoning							
	Inferences	.039	.046	z_WM	.215	2.518	.013
	Truth Table	.178	.131	z_WM	.194	2.027	.045
			.035	kbit	.214	2.307	.023
			.031	Stroop_inh	.185	2.231	.027
Syllogistic Reasoning							
	Syllogisms	.115	.097	z_WM	.253	2.944	.004
			.032	Stroop_inh	.188	2.182	.031
	Modal	.115	.121	z_WM	.348	4.254	<.001
Probabilistic reasoning		.022	.029	kbit	.171	1.982	.050

of the syllogistic inference performance for grade 2 students, adjusted R^2 = .115, supporting the importance of the WM development in syllogistic task even for single-model problem (we have excluded the two multiple-model problems in the regression analysis). In this kind of syllogistic construction task, inhibition of the knowledge about the terms (the semantics) in the syllogisms (e.g. for sure that not all the skiers are vegetarians in the real world) is essential for the successful solving of the syllogism. The inhibitory control ability of children develops later than the WM and is commonly thought to be increased with the WM. See Table 8 for the details of the results.

For both grades, adjusted R^2 of the significant model were very small for probabilistic reasoning tasks, adjusted $R^2 = .33$ and adjusted $R^2 = .22$ for grade 1 and grade 2 students respectively. The factors that entered the model were not the same, WM composite for grade 1 students but KBIT for grade 2 students. It seems that the factors affect the performance of probabilistic reasoning were a bit more complicated and different from the other reasoning tasks.

5.4 Discussions

Previous studies have found diverse results of the effects of executive functions and education in different reasoning tasks, especially for reasoning tasks with the possibility of employing heuristic strategies, in children, preadolescents or adolescents. Reasoning is a very high level of cognitive processes. As expected, different type of reasoning tasks require different cognitive processes (or at least the degree of involvement of different cognitive capacities and abilities are different). The effect of individual differences such as cognitive development, reasoning styles, level of education and effort to solve the problem can change the performance in different tasks, especially during the stage of rapid development and “mindware” change. The importance of different executive processes on different reasoning changes during the period of the preadolescence. This may due to the development of such capacities and abilities which facilitate participants to employ different processes/strategy to solve the problems (such as from system 1 to system 2). Different from Santamaría et al. (2013), which found significant effect of the verbal working memory on reasoning performances for both graders, working memory was the significant predictor for reasoning performances in the second graders, but not first graders.

With the development of the metacognition ability, and after being able to detect the existence of alternatives, the importance of the available/amount of cognitive resources starts to pay a more important role. Our results suggested that WM is important for the reasoning of even single-model problems (the syllogistic construction problem) in children aged 11 to 13, which seems less important in adults for solving easy problem.

Problem difficulty

Regarding our first prediction, we found results as what the MMT suggested: the three-model problems are significantly more difficult than the two-model problems and single-model problems. Participants had a very low percentage of correct responses (>= 28%) for all the three-model problems. Syllogisms reminded rather difficult for our participants, especially the two multiple-model problems. For the two-model problem, participants had to reject the automatic processing output by matching/atmosphere heuristic, the “Some … are …” conclusion, and give the “no valid conclusion” response. The construction of two mental models and searching for counterexamples (the fleshing out process) are essential in this case. While for the three-model problem, participants have to construct three mental models and as the conclusion has a different quantifier (Some … are not …) from the two premises, the use of the atmosphere/matching heuristic is not possible. They are difficult even for adults. The percentage of correct response was zero for this problem.

Metalogical tasks versus inference tasks

In general, we found improvements for the metalogical tasks, except the two problems with negation in the evaluation of propositional attitudes task and the second problem of the modal syllogistic task (which is single-model). We found similar percentages of correct responses for the modus tollens (MT) and denial of antecedent (DA) problems in the propositional inference task and the syllogistic inference problems as previous studies (e.g. Klaczynski, Schuneman, & Daniel, 2004). Our participants performed much better in the MT problem than the DA problem, in line with hypothesis of the mental model theory.

Effect of individual differences the grade

The executive function capacities develop with age in children around the ages of our participants. The improvement of the reasoning tasks may due to both the increase in executive function capacities and the experiences and formal training gain with the grade level. According to our regression results, it seemed that the increase of WM capacities is more important for the improvement of the reasoning performances, especially for grade 2 students. More specifically, for the single-model syllogistic problem, the development of WM and executive functions (as measured by the Stroop task) supported the view that such developments in the respective cognitive resources might be necessary to facilitate children's ability to inhibit the irrelevant background knowledge and engage in a context-free formal reasoning processes. We found more involvement of the fluid intelligence for first graders but the WM for second graders. Again, we found evidence supporting the MMT and dual-process theory's view that development of the WM is an important factor of reasoning performances in pre-adults. The fluid intelligence remains as the most powerful predictor of the reasoning performances, especially for grade 1 students for the propositional inference task and the modal syllogistic task.

Reference

Barrouillet, P. (2011). Dual-process theories of reasoning: The test of development. *Developmental Review, 31*(2), 151-179.

Barrouillet, P., Gauffroy, C., & Lecas, J. F. (2008). Mental models and the suppositional account of conditionals. *Psychological Review, 115*(3), 760.

Barrouillet, P., Grosset, N., & Lecas, J. F. (2000). Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition, 75*(3), 237-266.

Barrouillet, P., & Lecas, J. F. (1998). How can mental models theory account for content effects in conditional reasoning? A developmental perspective. *Cognition, 67*(3), 209-253.

Barrouillet, P., & Lecas, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning, 5*(4), 289-302.

Barrouillet, P., Markovits, H., & Quinn, S. (2002). Developmental and content effects in reasoning with causal conditionals. *Journal of Experimental Child Psychology, 81*(3), 235-248.

Buehner, M., Krumm, S., & Pick, M. (2005). Reasoning= working memory ≠ attention. *Intelligence, 33*(3), 251-272.

Chiesi, F., Primi, C., & Morsanyi, K. (2011). Developmental changes in probabilistic reasoning: The role of cognitive capacity, instructions, thinking styles, and relevant knowledge. *Thinking & Reasoning, 17*(3), 315-350.

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain.

Evans, J. S. B. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*(3), 378-395.

Evans, J. S. B. (2007). On the resolution of conflict in dual process theories of reasoning.

Thinking & Reasoning, 13(4), 321-339.

Evans, J. S. B. (2010). Intuition and reasoning: A dual-process perspective. *Psychological Inquiry, 21*(4), 313-326.

Evans, J. S. B., & Over, D. E. (1996). Rationality in the selection task: Epistemic utility versus uncertainty reduction.

Evans, J. S. B., & Over, D. E. (2004). *If*: Oxford: Oxford University Press.

Evans, J. S. B., & Stanovich, K. E. (2013a). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science, 8*(3), 223-241.

García-Madruga, J. A., Elosúa, M. R., Gil, L., Gómez-Veiga, I., Vila, J. Ó., Orjales, I., . . . Duque, G. (2013). Reading comprehension and working memory's executive processes: An intervention study in primary school students. *Reading Research Quarterly, 48*(2), 155-174.

García-Madruga, J. A., Santamaría, C., Moreno-Ríos, S., Vila, Gómez-Chacón, Orenes, & Navarrete. (2009). *Prueba de razonamiento para secundaria (Reasoning test for secondary school).*

Gauffroy, C., & Barrouillet, P. (2009). Heuristic and analytic processes in mental models for conditionals: An integrative developmental theory. *Developmental Review, 29*(4), 249-282.

Gauffroy, C., & Barrouillet, P. (2011). The primacy of thinking about possibilities in the development of reasoning. *Developmental Psychology, 47*(4), 1000.

Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. S. B. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning, 10*(2), 175-195.

Inhelder, B., & Piaget, J. (1955). *De la logique de l'enfant à la logique de l'adolescent*. Paris: Presses Univ. de France. (2e éd. 1970.).

Inhelder, B., & Piaget, J. (1964). *Classification and seriation*: Routledge & Kegan Paul.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review, 109*(4), 646.

Johnson-Laird, P. N., Oakhill, J., & Bull, D. (1986). Children's syllogistic reasoning. *The Quarterly Journal of Experimental Psychology, 38*(1), 35-58.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *J Exp Psych*, 1-17.

Klaczynski, P. A. (2004). A dual-process model of adolescent development: Implications for decision making, reasoning, and identity. *Advances in child development and behavior, 32*, 73-123.

Klaczynski, P. A. (2005). Metacognition and cognitive variability: A dual-process model of decision making and its development. *The development of judgment and decision making in children and adolescents*, 39-76.

Klaczynski, P. A., Schuneman, M. J., & Daniel, D. B. (2004). Theories of conditional reasoning: a developmental examination of competing hypotheses. *Developmental Psychology, 40*(4), 559.

Kuhn, D., & Franklin, S. (2006). *The second decade: What develops (and how)*: Wiley Online Library.

Markovits, H. (2004). The development of deductive reasoning. *The nature of reasoning*, 313-338.

Markovits, H., & Barrouillet, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review, 22*(1), 5-36.

Markovits, H., Schleifer, M., & Fortier, L. (1989). Development of elementary deductive reasoning in young children. *Developmental Psychology, 25*(5), 787.

Markovits, H., & Vachon, R. (1989). Reasoning with contrary-to-fact propositions. *Journal of Experimental Child Psychology, 47*(3), 398-412.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*: Cambridge University Press.

Moshman, D. (1990). Rationality as a goal of education. *Educational Psychology Review, 2*(4), 335-364.

Moshman, D. (2004). From inference to reasoning: The construction of rationality. *Thinking & Reasoning, 10*(2), 221-239.

Moshman, D., & Franks, B. A. (1986). Development of the concept of inferential validity. *Child development*, 153-165.

Piaget, J. (1953). The origins of intelligence in children. *Journal of Consulting Psychology, 17*(6), 467.

Piaget, J. (1970). Science of education and the psychology of the child. Trans. D. Coltman.

Santamaría, C., Tse, P. P., Moreno-Ríos, S., & García-Madruga, J. A. (2013). Deductive reasoning and metalogical knowledge in preadolescence: a mental model appraisal. *Journal of Cognitive Psychology, 25*(2), 192-200.

Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought*: Yale University Press.

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence, 30*(3), 261-288.

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. *two minds: dual processes and beyond. Oxford University Press, Oxford*.

Thompson, V. A. (2010). Towards a metacognitive dual process theory of conditional reasoning. *Cognition and conditionals: Probability and logic in human thinking*, 335-354.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1130. doi:10.1126/science.185.4157.1124

Chapter 6

Generation and evaluation tasks in syllogistic reasoning: The role of working memory

Author: Tse, P.P., Madruga, J., Moreno-Ríos, S., & Bajo, M.T

Abstract: The mental model theory of reasoning claims that people construct abstract mental models when they solve reasoning problems. Johnson-Laird (1983) proposed that the working memory (WM) is important for solving reasoning problems as the mental models are constructed and manipulated within WM. Various studies have suggested that people perform forward reasoning in conclusion generation tasks (premises-driven) and backward reasoning in conclusion evaluation tasks (conclusion-driven). In this study, the role of WM was investigated in an experiment in which participants carried out first a conclusion generation and then an evaluation syllogistic task. It was found that WM capacity, particularly the visuospatial working memory (VSWM), correlated with the generation task but not with the evaluation task, suggesting that people might use different strategies according to the demands of the task. The effect of number of models of the syllogism in the generation task and the influence of atmosphere/matching heuristics, particularly in the evaluation task, were also confirmed. The results suggest a dual-process explanation of syllogistic reasoning, in which participants may use some non-logical heuristics as well as analytic strategies.

Keywords: syllogistic reasoning, working memory, atmosphere hypothesis, mental model theory

6.1 Research background

Ever since the first studies and definition by Aristotle, (categorical) syllogism has been one of the most widely studied topics in reasoning research. There are still many unsolved "mysteries" about syllogistic reasoning, such as the reasoning processes of different individuals and the heuristics and strategies used by reasoners. Syllogisms are logical arguments consisting of two premises and a conclusion. Each of the three statements describes the relationship between two terms with four possible quantifiers, all (A), some (I), none (E) and some not (O). The two terms in the conclusion are called the end terms and each appears in one of the premises. The middle (linking) term is in both premises and reasoners have to deduce the relationship between the two end terms in the conclusion by their relationships with the middle term, assuming that the two premises are always true. For example:

	Abstract	Concrete
Premise 1:	All A are B	All dogs are mammals
Premise 2 :	No B are C	No mammals are reptiles
-------------	-------------------------	--------------------------------
Conclusion:	No A are C	No dogs are reptiles

"A/dogs" and "C/reptiles" are the two end terms while "B/mammals" is the middle term. There are 256 possible syllogisms, of which only 27 are valid (the conclusion follows logically from the premises, assuming the premises are true). The term 'mood' is used to refer to the different combinations of quantifiers of the premises and conclusion. Therefore, the above syllogism has the mood AE-E. In addition, there are four possible arrangements of the terms, called 'figures', according to the order of the terms A, B and C in the premises:

Figure 1	Figure 2	Figure 3	Figure 4
A-B	B-A	A-B	B-A
B-C	C-B	C-B	B-C

There is an effect of the figure in the conclusion generated by reasoners, a trend in syllogisms in Figure 1 to A-C conclusions and in Figure 2 to C-A conclusions (see, for instance, Dickstein, 1978; Espino, Santamaría, & García-Madruga, 2000; García-Madruga, 1982; Johnson-Laird & Bara, 1984)

Syllogistic reasoning has been widely studied by researchers in the last five decades. There are a number of diverse theories and different theoretical views on the cognitive processes that underlie reasoning. For example, the atmosphere hypothesis and the matching hypothesis are two of the old, superficial, non-logical hypotheses about human reasoning. More importantly, the three (recent) major schools of logical explanations are the rules-based (mental logic) approach (e.g., Braine, 1978; Rips, 1994), the mental models theory (e.g., Bucciarelli & Johnson-Laird, 1999; Johnson-Laird, 1983, 2001; Johnson-Laird & Byrne, 1991) and the dual-processing theories (e.g., Evans, 1984; Evans, 2003, 2008; Evans & Over, 1996; Evans, 2006; see Khemlani & Johnson-Laird, for an overview of different syllogistic theories) for a complete discussion of different theories of syllogisms). In this article, we will focus on the reasoning processes of reasoners in solving syllogistic conclusion generation and evaluation tasks based on four main hypotheses/theories: atmosphere hypothesis, mental rules, mental models and dual-processing theories. Evaluation and generation tasks differ on whether the conclusion is generated by participants (generation task) or is presented to participants with instructions to evaluate its validity (evaluation task). As we will discuss later, the contrast between the two tasks permits us to evaluate predictions from different reasoning hypotheses. We also tested the importance of working memory (WM) capacities in reasoning (of the more difficult reasoning problems) as suggested by some theories, such as the mental model and dual-processing theories and also Rips' PSYCOP model. In the following paragraphs, we will review firstly the main assumptions of reasoning hypotheses/theories, and then the differences between evaluation and generation tasks and the possible role of WM as related to the assumptions of the theories.

Reasoning Theories

Atmosphere-matching Theory

The atmosphere hypothesis (Revlis, 1975; Sells, 1936; Woodworth & Sells, 1935); (see also, Begg & Denny, 1969) claims that people solve syllogisms with non-logical heuristics that do not necessarily lead to a logical valid conclusion. People tend to accept the conclusion which satisfies the following two principles:

1. Principle of quality: if there is at least one negative premise ('No' or 'Some...are not'), a negative conclusion is favoured; otherwise, a positive conclusion ('All' or 'Some').
2. Principle of quantity: if there is at least one particular premise ('Some' or 'Some...are not'), a particular conclusion is favoured; otherwise, a universal conclusion ('All' or 'None').

The matching strategy is a modified version of the atmosphere hypothesis suggesting that people choose the conclusion which matches the quantifier of the more 'conservative' premise, the premise representing a lower number of entities, favouring the particular over the universal and negative rather than affirmative, i.e. E > O = I > > A (e.g., García-Madruga, 1983; Wetherick, 1989; Wetherick & Gilhooly, 1995). The matching hypothesis is sometimes tested in the evaluation task (e.g., Stupple, Ball, & Ellis, 2013; Stupple & Waterhouse, 2009), though the classical studies were construction tasks.

Thus, in the present study, in order to explore the use of heuristics, such as atmosphere, we manipulated the congruency of the quantifiers of invalid conclusions in the evaluation task according to the atmosphere hypothesis. Half of the invalid conclusions were congruent with the atmosphere hypothesis while the other half were not. If participants used the atmosphere as a heuristic for solving the syllogistic problems, they would erroneously accept more invalid conclusions congruent with atmosphere than invalid incongruent conclusions. Problems with invalid incongruent-with-atmosphere conclusions (invalid-NoAtm) were expected to be easier than those with invalid congruent-with-atmosphere conclusions (invalid-Atm) because for the latter problem, the correct response is to reject the favoured heuristic conclusion.

The rule-based approach

The rule-based approach suggests that reasoners use some kind of inherent mental deduction rules to reach a conclusion. For example, Rips' deduction-system hypothesis, which is based on the computational implementation of the PSYCOP model (Rips, 1994), contains a set of forward and backward inference rules. Forward rules draw implications from premises and generate sets of new sentences/assertions to a conclusion. Backward rules, however, reason back from the conclusion (or a tentative conclusion) to find essential assertions for the argument. These rules help reasoners construct mental proofs in WM. Although the theory cannot easily explain the systematic errors committed by reasoners or the effects caused by the contents (e.g. belief bias effect) because the rules are syntactic in nature, it suggests the possibility of using backward rules in the evaluation task. Thus, forward rules are predicted to be used in the generation task, whereas backward rules are predicted in the evaluation task. In the present experiment, we will explore these predictions by comparing results from the two tasks.

Mental Model Theory

Contrary to the rule-based theory's suggestion that reasoning is syntactic in nature, the mental model theory (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991) proposes that reasoning is based on the semantic processes of building and manipulating models of the contents of premises. Reasoners form mental models of the premises to figure out the conclusion. There are three stages in the reasoning process: premise interpretation, premise integration and conclusion validation. In the first stage, both the terms and their categorical relations are converted to abstract tokens (abstract concepts like negation or quantifiers are represented as well). Reasoners then form provisional integrated initial mental model(s) of the syllogism in the second stage. Mental models are assumed to be constructed and manipulated in WM. Some people may construct an image of what a model represents from a certain point of view (Knauff & Johnson-Laird, 2002), which implies that mental models may be spatial (and iconic) in nature and draw on the resources of the visuospatial sketchpad (VSSP) of WM particularly. It is also assumed that a major component of reasoning is nonverbal. In order not to overload the WM, reasoners try to present as little information as possible in the mental models and thus represent only what is true according to the premises, and not what is false (Johnson-Laird & Byrne, 2002; Johnson-Laird et al., 2004), and some linguistic properties of the premises may not be represented. The third stage is the only real deduction process. It involves a model-based search for counterexamples to the initial model. Error in performance is mainly due to insufficient search for the counterexamples or mistakes in the construction of the models. Most of the syllogisms are multiple-model problems. The problem is more difficult if more models have to be constructed as reasoners have to maintain and consider all of them to justify the validity of the conclusion. This process loads on cognitive resources and takes time (Bell & Johnson-Laird, 1998; Copeland & Radvansky, 2004) and thus leads to errors and inefficiency in reasoning when reasoners have limited WM capacity to represent and manipulate mental models (Johnson-Laird

& Byrne, 1991) and/or limited time to search for counterexamples. Johnson-Laird and his colleagues have found results from many studies congruent with the predictions of mental model theory (e.g., Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991).

In the present study, different single and multiple-model problems were used to explore the effect of the number of mental models on WM. We predicted that in the conclusion generation task, multiple-model problems would put a higher load on WM, especially on the central executive, than single-model problems. However, such an effect might diminish in the conclusion evaluation task if the reasoning process were conclusion-driven by means of a backward reasoning process, as some rule theorists have proposed (Rips, 1983, 1994).

Dual Processing Theory

There is a long history of the hypothesis that human reasoning and the decision-making process consist of (at least) two different types of processing (e.g., Evans, 1984, 2003, 2006; Evans & Over, 1996; Evans & Stanovich, 2013b; García-Madruga, 1983; Neys, 2006; Wason & Evans, 1975). Type 1 processing is fast, automatic, heuristic and unconscious (Evans, 2008). The output is sometimes/often biased and acceptance of fallacies in logical arguments is common. Although it is more intuitive, it can produce fairly good results which are sufficient for day-to-day problem solving (Evans, 2007). The heuristic processes described in the atmosphere/matching hypothesis belong to Type 1 processing. Type 2 processing is thought to be slower, sequential, analytical, conscious, controlled and demanding on cognitive resources. Type 1 processing solves problems based on prior knowledge, beliefs and heuristics; while Type 2 processing solves problems based on logical means with the inhibition of knowledge-based biases (cognitive decoupling) and thus places heavy demands on cognitive resources. As the inhibition of Type 1 processing and the computation of Type 2 processing draw on executive

WM resources, people with higher WM spans, especially the central executive (CE), would perform better in "difficult" tasks, for example, in the condition where there is a conflict between belief and logic; and outputs of heuristics and logical processes.

There is still no common consensus on how the two systems work together: whether they are parallel-competitive (with a mechanism to solve the conflicts, Sloman, 1996) or default-intervention (Type 1 processing as default and Type 2 processing may or may not intervene (see also, Evans & Stanovich, 2013a for a review). Several researchers have proposed a control system or mechanism for the shifting from Type 1 to Type 2 processing in possible conflict resolution (Evans, 2009; Stanovich et al., 2011). They mainly proposed a tripartite structure, a System 3, which deactivates Type 1 processing. However, the mechanism of this shift is still under debate (but see the "feeling of rightness hypothesis" of Thompson, 2010).

One recent view is that there is no universal processing mechanism but that reasoners choose different strategies according to the task's demands (task features and available time), instruction, motivation, their own cognitive capacities, experience and training (the heuristic-analytic theory of reasoning: e.g., Evans, 2006). De Neys (2006) showed that different reasoning tasks that were expected to employ Type 2 processing were more time-consuming and demanding of cognitive resources, which in turn supported different reasoning processes, and were employed in different reasoning tasks according to the task's demands and the experimental setup. In the present experiment, we created conditions where the outputs of the two systems were in conflict (in the conclusion evaluation task) to explore the role of task demands and individuals' WM capacity in the resolution of this conflict.

Generation Tasks versus Evaluation Tasks

There is agreement in the literature of categorical syllogisms that reasoners' performance in conclusion generation tasks and evaluation tasks are different, for example, in studies on figural effect (e.g., Evans, Handley, Harper, & Johnson-Laird, 1999; Johnson-Laird & Bara, 1984; Quayle & Ball, 2000) and belief bias effect (e.g., Morley et al., 2004).

(Morley et al., 2004) suggested that the reasoning process in conclusion generation tasks is more likely to be premise-driven processing (forward reasoning, from the premises to the conclusion) while that in conclusion evaluation tasks is more likely to be conclusion-driven, backward reasoning. Also, the provision of the conclusion helps in the construction of the mental model(s) of the premises (e.g., Evans et al., 2001; Hardman & Payne, 1995; Klauer, Musch, & Naumer, 2000). This hypothesis is supported by previous studies, (usually) finding a figural effect of the premises in the generation task but not in the evaluation task; and a stronger belief bias effect in conclusion generation than in evaluation tasks. The mental model theory of categorical syllogisms is mainly based on generation tasks: the construction and evaluation of the conclusion is essentially premises-driven. Therefore, the backward reasoning hypothesis is not in line with the standard mental model theory.

Working Memory and Reasoning

WM is commonly thought to consist of 4 subsystems: the phonological loop, visuospatial scratchpad (VSSP), episodic buffer and central executive (Baddeley, 1997; Baddeley & Hitch, 1974). The first two subsystems are the two slave systems responsible for the retention and storage of verbal and visuo-spatial materials, respectively. The episodic buffer is in charge of the connection with long term memory. The central executive (CE) helps in the allocation of attention to relevant information, the inhibition of irrelevant information and inappropriate mental operations, the coordination of attention resources when more than one task must be done simultaneously and the updating of representations. As mentioned before, diverse theories of reasoning suggest the involvement of different components of WM in certain phases of the reasoning process. The presentation modality of the premises (and conclusions) draws on WM if the premises are not always online (available) when the premises are presented verbally (Capon et al., 2003).

Many different studies have found evidence of the involvement of WM in deductive inferences and particularly in syllogistic reasoning, especially the CE (e.g., Copeland & Radvansky, 2004; García-Madruga et al., 2007). Gilhooly and his collaborators have found significant correlations of reasoning performances with different WM measures in different syllogistic tasks, including conclusion generation tasks (e.g., Bacon, Handley, Dennis, & Newstead, 2008) and evaluation tasks (e.g., Gilhooly et al., 2002b) and with dual tasks design (Gilhooly et al., 1993; Gilhooly et al., 1999). Significant results were also found with both abstract (e.g., Gilhooly et al., 1993) and concrete materials (e.g., Bacon et al., 2008) and with different "types" of participants, e.g. grouping the participants according to their reasoning skills (Gilhooly et al., 1999) and reasoning strategies (Bacon et al., 2008) and whether or not they received pre-experiment training (Gilhooly et al., 1999).

We used the two visuospatial WM tasks in Shah and Miyake (1996), the arrow task and the letter task, to investigate the manner of involvement of the visuospatial WM in our two syllogistic tasks. The arrow task is a simple visuospatial WM task, which loads on the storage component of the visuospatial WM; while the complex letter task draws on both the processing and the storage components of the visuospatial WM simultaneously, as well as on the CE. (Capon et al., 2003) found that the performance of different syllogistic tasks correlated differently with the simple and complex tasks of verbal and visual WM and that visual presentation of the syllogism lowered the cognitive demand; also that the performance of participants improved. However, in another study by Bacon et al. (2008), performance was found to correlate with the simple spatial task but not with the complex spatial task with visually presented online premises. As one of our aims was to investigate the effect of task demand on WM, we opted to present the two tasks visually in order to eliminate the effect of presentation modality on cognitive resources. In addition, we used the operation span task (Turner & Engle, 1989), which loads mainly on the CE, and the reading span task (Daneman & Carpenter, 1980), which loads mainly on the verbal WM but also on the CE (Whitney, Arnett, Driver, & Budd, 2001). We predicted a higher involvement of the WM capacities in the performance of the generation task than in the evaluation task, in particular, a higher correlation with the measures of the complex tasks that load on the CE (the letter task, reading span task and operation span task).

Aims and predictions of the Study

In this study, we explored the importance of two factors, namely, the number of mental models of the syllogism and the capacities of different components of the WM of reasoners in two syllogistic tasks: generation and evaluation. If the cognitive processes involved in these two tasks are different, the results of the single and multiple-model syllogisms will be different. We also studied the use of the atmosphere heuristic in the evaluation task. We predicted the following:

Firstly, as mentioned before, the evaluation task was expected to be easier than the generation task. We predicted an overall better performance in the evaluation task than in the generation task. Secondly, we predicted that in the generation task, the single-model syllogisms would have more correct responses than multiple-model syllogisms, as participants had to construct and also search for the counterexample(s) of more mental models for the latter. Thirdly, for the evaluation task, we expected that the importance of the number of mental models would diminish. We manipulated the invalid conclusions so that the use of the atmosphere heuristic was favoured. If the participants did not construct mental models but employed the atmosphere heuristic to solve the problems, the number of mental models of the problems would no longer predict accuracy in solving the syllogisms. Participants could also reason backwards when the conclusion was provided.

Fourthly, in the evaluation task, invalid syllogisms were expected to have more *incorrect* responses, as participants generally exhibit a tendency to accept the conclusions without considering their actual validity (e.g., Klauer et al., 2000). Problems with invalid congruent-with-atmosphere conclusions (invalid-Atm) would be the most difficult among all the problem types as there is a conflict between the output of the atmosphere heuristic and the actual validity of the syllogism. Participants had to inhibit the use of the atmosphere heuristic and solve the problems analytically. We predicted a very low accuracy for this type of problem. Problems with invalid incongruent-with-atmosphere conclusions (invalid-NoAtm) would be easier than invalid-Atm problems as there is no conflict between the actual validity and the output of the atmosphere heuristic, i.e. rejecting the conclusion.

Fifthly, the performance of solving the multiple-model problems would have a higher correlation with WM measures, especially the CE loaded tasks (letter task, reading span task and operation span task). We expected less involvement of WM in the evaluation task, as the use of a heuristic strategy was encouraged.

6.2 Method

Participants

A total of 80 undergraduates from the University of Granada took part in the experiment. They were 59 females and 21 males with ages ranging from 18 to 39 years (mean = 23, SD = 4.3). No participants had formal training in logic.

Procedure

There were two syllogistic tasks and four WM tasks. The participants always performed the conclusion generation task first, then the conclusion evaluation task and finally the four WM tasks. The sequence of WM tasks was counterbalanced.

Tasks and Materials

Syllogistic Tasks

Syllogistic Material. We used the twenty-seven syllogisms that have a (natural) valid conclusion in this study. Five of them were used in the conclusion generation task while the remaining twenty-two problems were included in the evaluation task, such that each participant solved the same problem once only. The problems were counterbalanced in four versions of the task so that all the twenty-seven syllogisms were included in the evaluation task. The problems were classified into single and multiple-model, following (Johnson-Laird & Byrne, 1991).

Syllogistic conclusion generation task. Each participant had to state the conclusion of 5 syllogisms, with two single-model and three multiple-model problems. The terms in the syllogisms were professions or sports occupations (e.g. All bellboys are translators) in Spanish. Participants had to think aloud and provide concurrent written protocols detailing their working out. There were 4 random versions in this task. All the syllogisms had a valid conclusion. See

Table 9 for all the syllogisms used in the generation task.

Syllogistic conclusion evaluation task. Participants had to evaluate the validity of the given conclusions of the remaining 22 syllogisms (either valid or invalid). The problems were presented in Spanish. There were four random versions. In each version, half of the conclusions

Table 9

The 5 Syllogisms in the Generation Task

Version	2 single-model problems		3 multiple-model problems		
1 and 2	AI2 – I(CA)	IA4 – I(AC)	OA3 – O(AC)	IE2 – O(AC)	EI2 – O(CA)
3 and 4	IA1 – I(AC)	AI4 – I(CA)	AO3 – O(CA)	EI1 – O(CA)	IE1 – O(AC)

were valid and the other half were invalid (eleven valid and eleven invalid). Half of the invalid conclusions did not agree with the atmosphere hypothesis (NoAtm) and half of them agreed with the atmosphere hypothesis (Atm) (five NoAtm and six Atm in two random versions; and six NoAtm and five Atm in the other two random versions). The problems were not repeated in the two syllogistic tasks (see Appendix D for the list of syllogisms in the two tasks and the percentages of correct responses).

Two practice trials were presented before the 22 experimental trials. The standard deductive reasoning instruction was used. Participants were asked to assume that the two premises were always true and the conclusion was valid only if it followed logically from the two premises. In the syllogistic evaluation reasoning task, each premise was presented consecutively for 3 seconds followed by 500ms of a blank page. The word “therefore” was then displayed. Participants had to press a key when they were ready to read the conclusion. The conclusion was then displayed until the participant pressed a key to indicate his/her response. There was a time limit of 10 seconds for the response. The time between the “Therefore” signal and participant’s response indicating that they were ready to read the conclusion, and the time after that until they indicated their response were measured.

Working Memory tasks

Simple visuospatial task (arrow task). Each test item was an arrow pointing in one of the eight possible orientations (0º, 45º, 90º, 135º, 180º, 225º, 270º, 315º and 360º). The stimuli were presented in sets of ascending size from two to six items, with **three sets** at each level. There were fifteen sets in total. Each arrow remained on the screen for 1000 ms, with an inter-stimulus interval of 250 ms. One point was awarded for each correct arrow in each fully correct set. The maximum score for this task was 60.

Complex visuospatial task (letter task). Each item was either the normal or mirror image of one of the five letters F, J, L, P and R in one of the seven possible orientations (45º, 90º, 135º, 180º, 225º, 270º, 315º and 360º), except the upright position. Participants were informed of the five letters, that they were capital letter and that no “b” would be presented. Test items were presented in sets of ascending size, from two to five items, with five sets at each level, twenty sets in total. Presentation was constrained so that opposing orientations were not presented successively within a set and that each orientation appeared only once per set. The letters remained on the screen for a maximum of 2200 ms and participants were asked to respond whether the image was a normal (by pressing “P”) or mirror image (by pressing “Q”) of the letter when it was presented. After each set, a screen with eight numbers placed in the eight positions was displayed (see Figure 1 for an example).

Participants had to indicate the number corresponding to the position of the top of each letter in the order that the letters were presented. For example, "ꓤ" is a normal "R" with its top pointing downward. Participants had to press "P" when they saw the image "ꓤ" and then press the number corresponding to the downward position, which was "7" (see Figure 1) in this case, after the whole letter set was displayed in its presentation order. The numbers changed after each set so that participants had to remember the position rather than the number. One point per letter was awarded for each correct set. Accuracy of the normal/mirror image judgment was also recorded.

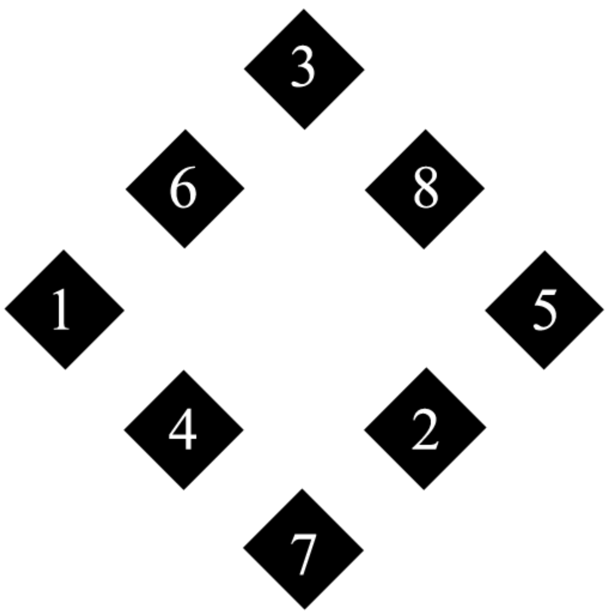


Figure 1. An example of the eight numbers representing the position of the top of the letter.

Reading span task. We used the Spanish version of the reading span task of Daneman and Carpenter (1980), which aimed to measure the verbal WM capacity of the participant, although CE is also involved to a lesser extent. Participants had to read sentences aloud and then remember the last word of each sentence. Sentences were presented on a computer screen in sets of increasing size from two to six sentences. There were five sets in each level except for the six-sentence sets with three sets only (eighty-eight sentences in total). Participants were asked to recall the last words in correct order after the presentation of each set. One point was awarded for each word if the participant reported the whole set of words in correct order.

Operation span task. This task (adapted from Turner & Engle, 1989) is highly correlated with the CE capacity. In each trial, participants had to verify a mathematical operation and then memorise a monosyllabic word (in Spanish). Three sets of the levels from two to six words were presented in random order (fifteen sets in total). Participants had to report all the words in the correct order after each set. One point was recorded per word for each correct set.

6.3 Results

We used the complete set of the twenty-seven valid syllogisms. These syllogisms are not distributed in equal numbers between the conditions of interest. Therefore, specific predictions determine which subsets of syllogisms are to be used to test them. Following this basic criterion, we tested, for example, the set of ten syllogisms which were used in both tasks to evaluate our first hypothesis involving the task effect of the generation and evaluation tasks; for the predictions involving the atmosphere heuristics, we tested the eighteen multiple-model syllogisms[13].

Overall Comparison between the Two Tasks

In general, participants achieved a better performance for the evaluation task. We tested the same ten syllogisms (used in both tasks). The percentages of correct responses in the

[13] We used IBM SPSS Statistics 20 for all the statistical analyses.

Table 10

Percentages of Correct Responses of Each Syllogism in Generation Task

Syllogism	Conclusion	Mental model	Accuracy
IA-I1	I-AC	single	85%
AI-I2	I-CA	single	80%
IA-I4	I-AC	single	60%
AI-I4	I-CA	single	66%
	Total single-model		73%
IE-O1	O-AC	multiple	73%
EI-O1	O-CA	multiple	15%
IE-O2	O-AC	multiple	28%
EI-O2	O-CA	multiple	45%
OA-O3	O-AC	multiple	23%
AO-O3	O-CA	multiple	32%
	Total multiple-model		35%
	Total		51%

generation and evaluation tasks were 50.25% and 63.50%, respectively and the difference was significant, t(79) = 3.33, $p < .01$ (see Tables 10 and 11).

Effect of Number of Mental Models

Generation Task. The percentage of correct responses to the syllogisms of the generation task is shown in Table 10. Participants' performance on multiple-model syllogisms was reliably worse than that on single-model syllogisms, with the percentages of correct responses being 35% and 73%, respectively, t(79) = 7.58, $p < .01$.

Evaluation task. In the evaluation task, the overall average percentages of correct responses for single-model syllogisms and multiple-model syllogisms were 66% and 49%, respectively, t(79) = 6.42, $p < .01$ (see Table 11). For problems with valid conclusions, more correct responses were given for single (68%) than for multiple-model (61%) problems but the difference was only marginally significant, t(79) = 1.98, $p = .051$.

Table 11

Percentages of Correct Responses and Response Latencies in the Main Kind of Syllogistic Problem in the Evaluation Task

Conclusion type	Mental Model	Condition	Accuracy	Response Latency
Invalid		single-Atm	14%	3676
		single-NoAtm	69%	3641
	single-model		64%	3662
		multiple-Atm	31%	3929
		multiple-NoAtm	50%	3781
	multiple-model		37%	3879
	Total invalid		45%	3769
Valid		single-Atm	68%	3260
		single-NoAtm	---	---
	single-model		68%	3260
		multiple-Atm	71%	3782
		multiple-NoAtm	46%	4028
	multiple-model		61%	3874
	Total valid		63%	3653
	Total		54%	3679

We performed several t-tests comparing performances of the two tasks. For multiple-model syllogisms, the percentages of correct responses of the *same* syllogisms (the ten syllogisms that we used in both tasks) in the generation task and the evaluation task were 35% and 52%, respectively, and the difference was significant, t(79) = 3.66, $p < .01$ (see Table 12). We also tested the *different* syllogisms (the remaining twelve multiple-model syllogisms which were not used in the generation task) as control. The percentage of correct responses of the *different* syllogisms was 48% and the t-test showed that the performance of the 12 *different* syllogisms was better than that of the generation task, t(79) = 3.60, $p = .01$ (see Table 13). The t-tests suggested that participants had a higher percentage of correct responses in the evaluation task for both the *same* and *different* syllogisms for multiple-model problems. However, such a difference was absent for single-model *same* syllogisms, t(79) = 1.39, $p = .17$. In addition, to test if the improvement in the evaluation task was due to the training effect, we compared the percentages of correct responses of the first half of the trials with that of the second half in the evaluation task. The t-tests were *not* significant for the single-model problems, t(79) = .61, $p = .54$, nor the multiple-model problems, t(79) = .84, $p = .41$, nor the overall percentages of correct responses, t(79) = .23, $p = .82$.

In addition, we selected two sets of single and multiple-model problems for the effect of number of mental models in the evaluation task. There were seventeen single-model problems: nine valid-Atm and eight invalid-NoAtm syllogisms; and eighteen multiple-model problems: twelve valid-Atm and six invalid-NoAtm (see Appendix C, yellow versus blue). The difference between the single and multiple-model syllogisms tested was significant, t(79) = 2.09, $p = .04$. The latency results were consistent with the accuracy results. Response latency (filtered) of correct responses of single-model syllogisms was faster than that of multiple-model syllogisms

(3424ms and 3877ms respectively, t(79) = 4.19, $p < .01$).

Table 12

*Results of **the Same** Syllogisms in the Generation and Evaluation Tasks*

	generation task	evaluation task	t(79)	*p*
Single-model	73% (.36)	79% (.28)	1.39	.167
Multiple-model	35% (.28)	53% (.27)	3.66	< .001**
Overall	50% (.23)	64% (.22)	3.33	.010**

Note. Standard deviation in parentheses.

** $p < .01$; two-tailed.

Table 13

*Results of **Different** Syllogisms in the Generation and Evaluation Tasks*

	generation task	evaluation task	t-value	*p*
Single-model	73% (.36)	60% (.26)	2.53	.014**
Multiple-model	35% (.28)	48% (.15)	3.60	.001**
overall	50% (.23)	51% (.13)	0.40	.691

Note. Standard deviation in parentheses.

** $p < .01$; two-tailed.

Importance of the atmosphere effect.

We performed the test with only the multiple-model syllogisms, due to the lack of single-model valid-NoAtm syllogism. For valid multiple-model problems, the difference between valid-Atm (71%) and valid-NoAtm (46%) was significant, t(79) = 5.88, $p < .01$. For invalid multiple-model problems, the difference between invalid-Atm (31%) and invalid-NoAtm (50%) was also significant, t(79) = 4.08, $p < .01$.

To test whether the number of mental models is relevant in automatic processes (when the valid conclusion agrees with the output of the atmosphere hypothesis), we compared the single and multiple-model syllogisms of valid-Atm and invalid-NoAtm problems. The difference between single and multiple model syllogisms for valid-Atm problems was not significant (68% and 71%, respectively; t(79) = 1.10, $p = .28$); however, for invalid-NoAtm problems, the single model syllogisms were easier than multiple model ones (69% and 50%; t(79) = 4.43, $p < .01$).

Importance of WM in different syllogistic tasks

Generation task. Regarding the correlation analysis, the letter score correlated significantly with the overall percentage of correct responses and percentages of correct responses to both single and multiple-model problems. We calculated the WM composite by averaging the z-scores of the four WM measures. The WM composite correlated with the overall percentage of correct responses and the arrow score correlated significantly with the single-model problems only (see Table 14).

Table 14

Correlations Between the WM Measures and the Accuracy of the Generation Task

Mental Model	arrow score	letter score	operation score	reading score	WM composite
single-model	.25*	.26*	.08	.14	.19
multiple-model	.03	.32**	.14	.07	.16
Total	.18	.40**	.15	.14	.23*

* $p < .05$; ** $p < .01$; two-tailed.

Evaluation task. The overall percentage of correct responses did not correlate with any of the WM measures. The percentages of correct responses of invalid single-model problems correlated with the WM composite and arrow score. More detailed analysis showed that it was the single-model invalid-Atm problem which correlated with the arrow score and the single-model invalid-NoAtm problem which correlated with the WM composite (see Table 15).

6.4 Discussions

In some everyday situations, people have to evaluate whether a given conclusion is necessary from a set of premises (evaluation task). In other situations, they have to create a conclusion, given the same set of premises (generation task). We assume that some mental processes are different and some are similar in both kinds of situation. In this study, we tested how automatic heuristics influence these processes and how the components of WM could relate with them.

Three factors were manipulated in the construction of the syllogisms to identify the differences between reasoning processes in evaluation and generation tasks: the number of different valid situations consistent with the premises (the number of mental models), whether the conclusion was consistent with the “automatic” atmosphere heuristic and the validity of the conclusion (in evaluation tasks). In general, syllogisms of more mental models require greater involvement of (Type 2 processing and therefore) WM. Also, facilitation by automatic atmosphere heuristic will induce the acceptance of valid congruent conclusions (Type 1 processing), but Type 2 processing will be required to discard the congruent conclusion when it is invalid. If processing in the evaluation task is based on backward processes centered on the conclusion and in the generation task on forward processes centered on the premises, the effect of the number of mental models

Table 15

Correlations Between the WM Measures and the Percentages of Correct Responses in the Evaluation Task

conclusion type	Mental Model	condition	arrow score	letter score	operation score	reading score	WM composite
Invalid		single-Atm	.31**	.10	.05	-.06	.12
		single-NoAtm	.20	.12	.05	.16	.24*
	single-model		.25*	.15	.05	.10	.25*
		multiple-Atm	.07	.06	-.16	-.13	-.06
		multiple-NoAtm	-.03	.09	.06	.08	.07
	multiple-model		.03	.10	-.08	-.05	.00
	Total invalid		.18	.17	-.03	.02	.15
Valid		single-Atm	-.04	.03	.07	.05	.05
		single-NoAtm	--	--	--	--	--
	single-model		-.04	.03	.07	.05	.05
		multiple-Atm	.08	-.05	.21	.10	.10
		multiple-NoAtm	-.04	-.01	-.16	.00	-.15
	multiple-model		.03	-.04	.01	.06	-.05
	Total valid		.02	.00	.06	.09	.01
Total			.13	.11	.02	.11	.02

** $p < .01$ * $p < .05$; two-tailed.

and the atmosphere heuristic should be different in the two tasks. Therefore, the evaluation task is expected to be more sensitive to automatic processes (the atmosphere heuristic) that connect the conclusion with the premises with just the quantifiers. On the other hand, the generation task will be more sensitive to the number of alternative conclusions of the premises (number of mental models).

Participants may simply overlook the semantic content of the premises in the evaluation task (e.g., Evans, Barston, & Pollard, 1983; Evans et al., 2001) and apply some superficial strategies like matching/atmosphere or simply guessing to solve the problem (as supported by the absence of belief-bias effect in the literature), due to the task's demands (Hardman & Payne, 1995; Markovits & Nantel, 1989). More specifically, in the evaluation task, participants simply had to determine the validity of the conclusion; while in the generation task, they had to produce a conclusion from a number of logical possibilities by constructing the mental models and searching for counterexamples. The generation task, therefore, imposes a much heavier load on our limited WM, more specifically, for the multiple-model problems.

In line with the findings of previous studies, our first prediction that participants would generally perform better in conclusion evaluation tasks than in generation tasks was confirmed, more specifically for multiple-model syllogisms. As expected, we found that the percentage of correct responses of multiple-model syllogisms in the evaluation task was larger than that in the generation task. This task effect was present when comparing the syllogisms used in both tasks (same syllogisms) but also syllogisms that were used only in the evaluation task with those in the generation task (different syllogisms, see Tables 12 and 13)[1].

Regarding the results of WM, as expected, we found significant correlations of the WM measures, especially the letter score, with the percentages of correct responses of single and multiple-model syllogisms and the overall results in the generation task. We found a higher correlation of the multiple-model problems than the single-model problems with the visuospatial WM measure. However, unlike some previous studies, we did *no*t find significant correlations with the reading span score or the operation span score in either task. This may be due to different experimental settings between our study and other studies and to individual differences. In our study, percentages of correct responses in the evaluation task did not correlate with the WM measures (except for the invalid single-model problems). Besides the fact that the use of superficial strategies was encouraged, heuristic processes may often dominate over logical processes, especially when the task demand overloads the cognitive resources. Our evaluation task was especially difficult as the premises were offline. Participants had to hold the two premises in their WM during the whole deductive process/trial. This task design may overload the WM of participants and thus the use of heuristics was common. Gilhooly et al. (1993) found that when participants used heuristics, the visuospatial scratch pad (VSSP) of WM was not involved as these heuristics impose less of a load on WM.

Other results are consistent with the use of backward reasoning strategies in the evaluation task. In support of this interpretation, we found atmosphere effect in the evaluation task for both multiple-model valid and invalid conclusion problems. This means the percentages of correct responses of multiple-model valid-Atm and invalid-NoAtm problems were significantly higher than that of valid-NoAtm and invalid-Atm problems, respectively. It seemed that participants generally used heuristic or backward reasoning. Participants tend to accept the initial mental model as valid (even without searching for counterexamples) or use heuristics to evaluate the validity of the conclusion.

Similarly, for valid conclusions, we found significant differences between the percentages of correct responses of single and multiple-model syllogisms in the generation task but not (p = 0.51) in the evaluation task; likewise, in the evaluation task, single-model valid-Atm were *not* significantly more difficult than multiple-model valid-Atm problems. In the evaluation task, while the atmosphere of the conclusion was a factor in the percentage of correct responses, the number of mental models of the syllogisms did not have an effect for valid conclusions. This result again suggests that participants solved the syllogisms in the evaluation task by conclusion-driven processing and used the atmosphere of the syllogism as a superficial heuristic to solve the problem. In contrast, the number of mental models had an effect on the performance of invalid-conclusion problems in the evaluation task. The percentage of correct responses of the multiple-model invalid-NoAtm problems was significantly smaller than that of single-model invalid-NoAtm problems.

According to the mental model theory, people always construct the initial mental model(s). For single-model problems, the initial model agrees with the atmosphere/matching heuristic output and is sufficient to draw a valid conclusion. Accordingly, our results showed faster responses and a higher percentage of correct responses in this condition. However, for multiple-model problems, in addition to the initial model yielded by Type 1 processing, reasoners had to construct more mental models from analytic processes carried out by Type 2 processing and thus the response latency was longer and the probability of errors was higher, even when the reasoning process was conclusion-driven in the evaluation task. The logic is similar regarding the difference between the percentages of correct responses of single and multiple-model invalid-NoAtm problems. The heuristic output of this type of problem to reject the conclusion agrees with the correct response. It is especially easy for single-model problems, as the initial model gives the correct response, which agrees with the heuristic. However, for multiple-model problems, some participants may make mistakes during the three phases of model construction and counterexample searching and therefore it led to more *incorrect* responses. As expected, invalid-Atm problems have an especially low percentage of correct responses. The extra difficulty of invalid conclusions was mainly due to the invalid-Atm problems. Solving the invalid-Atm problems involves the inhibition of (the processing and output of) Type 1 processing and activation of Type 2 processing and thus is more challenging.

Interestingly, the single-model invalid-Atm and invalid-NoAtm problems correlated significantly with the arrow score and WM composite, respectively. The invalid-NoAtm problems were easy (69 % of correct responses). A majority of the participants may apply consciously explicit heuristics to reject these conclusions because the conclusions are not congruent with the atmosphere heuristic. The significant correlation may suggest that conscious application of superficial strategies may load on cognitive resources, or that participants with higher WM could solve it better.

Regarding the single-model invalid-Atm problem, there was only one such problem in the task:

"All B are A"

"All C are B"

"All A are C" (invalid-Atm conclusion provided, "**All C are A**" is valid).

It was difficult (only 14% of correct responses) regardless of the fact that it is a single-model problem. The difference between the valid and the given invalid conclusion is the order of the end terms; moreover, in this Figure 2 syllogism, the invalid "All A are C" is the conclusion yielded by the effect of the figure. Participants who used atmosphere heuristic blindly would accept the conclusion as valid, instead of giving the correct rejection response, as the atmosphere heuristic does not take the figure into account and is not sensitive to the asymmetry of affirmative universal premises in syllogisms. The significant correlation with the arrow score suggests that the visuospatial WM is important for such syllogisms in which the figure of the conclusion can determine its validity.

In summary, our results supported the following conclusions: In general, the difficulty of making inferences with syllogisms seems to depend on whether the task requires Type 2 processing or just Type 1 processing. There is a difference in terms of difficulty between single and multiple-model problems, as predicted by the mental model theory; likewise, WM plays a role in the analytical solving of multiple-model problems. Effect of non-logical superficial processes was also found to be consistent with the atmosphere (and matching hypothesis). The kind of inference task, generation or evaluation, induces different ways of solving syllogisms. Our results supported the view that people use different reasoning strategies as well as superficial heuristics according to the task's demands: conclusion generation tasks were more prone to premise-driven processing; while evaluation tasks were more prone to conclusion-driven processing. These results are consistent with the Rule-based PSYCOP model proposed by Rips (1994) in which reasoners can employ some backward reasoning strategies in evaluation tasks. The dual-processing approach can help us describe and analyse the cognitive processes underlying syllogistic reasoning and the role of WM in processing, especially in Type 2 processing. Further studies could be on individual differences in generation and evaluation tasks and the trigger and mechanism underlying the "switch" from Type 1 processing to Type 2 processing.

Reference

Bacon, A. M., Handley, S. J., Dennis, I., & Newstead, S. E. (2008). Reasoning strategies: The role of working memory and verbal-spatial ability. *European Journal of Cognitive Psychology, 20*(6), 1065-1086.

Baddeley, A. (1997). *Human memory: Theory and practice*: Psychology Press.

Baddeley, A., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation, 8*, 47-89.

Begg, I., & Denny, J. P. (1969). Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning errors. *Journal of Experimental Psychology: , Memory, and Cognition, 81*(2), 351-354.

Bell, V. A., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science, 22*(1), 25-51. doi:10.1016/S0364-0213(99)80034-2

Braine, M. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review, 85*(1), 1.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science, 23*(3), 247-303. doi:10.1016/S0364-0213(99)00008-7

Capon, A., Handley, S., & Dennis, I. (2003). Working memory and reasoning: An individual differences perspective. *Thinking & Reasoning, 9*(3), 203-244.

Copeland, D., & Radvansky, G. (2004). Working memory and syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 57*(8), 1437-1457. doi:10.1080/02724980343000846

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466.

De Neys, W. (2006). Automatic–heuristic and executive–analytic processing during reasoning: Chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology, 59*(6), 1070-1100.

Dickstein, L. S. (1978). The effect of figure on syllogistic reasoning. *Memory and Cognition, 6*(1), 76-83.

Espino, O., Santamaría, C., & García-Madruga, J. A. (2000). Figure and difficulty in syllogistic reasoning. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*.

Evans, J. S. B. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75*(4), 451-468.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10).

Evans, J. S. B. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*(3), 378-395.

Evans, J. S. B. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning, 13*(4), 321-339.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol., 59*, 255-278.

Evans, J. S. B. (2009). How many dual-process theories do we need? One, two, or many?

Evans, J. S. B., Barston, J., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition, 11*(3), 295-306. doi:10.3758/BF03196976

Evans, J. S. B., Handley, S. J., & Harper, C. N. J. (2001). Necessity, possibility and belief: A study of syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 54*(3), 935-958. doi:10.1080/713755983

Evans, J. S. B., Handley, S. J., Harper, C. N. J., & Johnson-Laird, P. N. (1999). Reasoning about necessity and possibility: A test of the mental model theoriy of deduction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1495-1513.

Evans, J. S. B., & Over, D. E. (1996). Rationality in the selection task: Epistemic utility versus uncertainty reduction.

Evans, J. S. B., & Stanovich, K. E. (2013a). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science, 8*(3), 223-241.

Evans, J. S. B., & Stanovich, K. E. (2013b). Theory and metatheory in the study of dual processing: Reply to comments. *Perspectives on psychological science, 8*(3), 263-271.

García-Madruga, J. A. (1982). Un estudio sobre el efecto de la figura en el razonamiento silogístico. *Estudios de Psicología*, *3*(11), 23-32.

García-Madruga, J. A. (1983). Un modelo general sobre el razonamiento silogístico: Doble procesamiento y fase de comprobación con verificación. *Revista de Psicología General y Aplicada*.

García-Madruga, J. A., Gutiérrez, F., Carriedo, N., Luzón, J. M., & Vila, J. O. (2007). Mental models in propositional reasoning and working memory's central executive. *Thinking & Reasoning, 13*(4), 370-393.

Gilhooly, K. J., Logie, R., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & Cognition, 21*(1), 115-124.

Gilhooly, K. J., Logie, R., & Wynn, V. (1999). Syllogistic reasoning tasks, working memory, and skill. *European Journal of Cognitive Psychology, 11*(4), 473-498.

Gilhooly, K. J., Logie, R., & Wynn, V. (2002b). Syllogistic reasoning tasks and working memory: Evidence from sequential presentation of premises. *Current Psychology, 21*(2), 111-120. doi:10.1007/s12144-002-1007-x

Hardman, D. K., & Payne, S. J. (1995). Problem difficulty and response format in syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 48*(4), 945-975.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*: Harvard University Press.

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences, 5*(10), 434-442.

Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition, 16*(1), 1-61.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*: Lawrence Erlbaum Associates, Inc.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review, 109*(4), 646.

Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111*(3), 640.

Khemlani, S., & Johnson-Laird, P. N. (2012). Theories of the Syllogism: A Meta-Analysis. *Psychological Bulletin, 138*(3), 427-457. doi:10.1037/a0026841

Klauer, K. C., Musch, J., & Naumer, B. (2000). On Belief Bias in Syllogistic Reasoning. *Psychological Review, 107*(4), 852-884. doi:10.1037/0033-295X.107.4.852

Knauff, M., & Johnson-Laird, P. N. (2002). Visual imagery can impede reasoning. *Memory & Cognition, 30*(3), 363-371.

Markovits, H., & Nantel, G. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition, 17*(1), 11-17.

Morley, N. J., Evans, J. S. B., & Handley, S. J. (2004). Belief bias and figural bias in syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 57*(4), 666-692.

doi:10.1080/02724980343000440

Neys, W. D. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological science, 17*(5), 428-433.

Quayle, J. D., & Ball, L. J. (2000). Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 53*(4), 1202-1223. doi:10.1080/713755945

Revlis, R. (1975). Two models of syllogistic inference: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior, 14*(2), 180-195.

Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review, 90*(1), 38.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*: MIT Press.

Sells, S. B. (1936). The atmosphere effect: an experimental study of reasoning. *Archives of Psychology (Columbia University).*

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: an individual differences approach. *Journal of Experimental Psychology: General, 125*(1), 4-27.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). The complexity of developmental predictions from dual process models. *Developmental Review, 31*(2), 103-118.

Stupple, E. J., Ball, L., & Ellis, D. (2013). Matching bias in syllogistic reasoning: Evidence for a dual-process account from response times and confidence ratings. *Thinking & Reasoning, 19*(1), 54-77. doi:10.1080/13546783.2012.735622

Stupple, E. J., & Waterhouse, E. F. (2009). Negations in syllogistic reasoning: Evidence for a heuristic–analytic conflict. *The Quarterly Journal of Experimental Psychology, 62*(8), 1533-1541. doi:10.1080/17470210902785674

Thompson, V. A. (2010). Towards a metacognitive dual process theory of conditional reasoning. *Cognition and conditionals: Probability and logic in human thinking*, 335-354.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127-154.

Wason, P. C., & Evans, J. S. B. (1975). Dual processes in reasoning? *Cognition, 3*(2), 141-154.

Wetherick, N. (1989). Psychology and syllogistic reasoning. *Philosophical Psychology, 2*(1), 111-124.

Wetherick, N. E., & Gilhooly, K. J. (1995). Atompshere, matching, and logic in syllogistic reasoning. *Current Psychology, 14*(iss.3), 169-178.

Whitney, P., Arnett, P. A., Driver, A., & Budd, D. (2001). Measuring central executive functioning: What's in a reading span? *Brain and Cognition, 45*(1), 1-14.

Woodworth, R. S., & Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology, 18*(4), 451-460.

Chapter 7

Inhibitory Mechanism of the Matching Heuristic in Syllogistic Reasoning

Reference: Tse, P.P., Moreno-Ríos, S., García-Madruga, J. & Bajo, M.T. (2014). Inhibitory mechanism of the matching heuristic in syllogistic reasoning. Acta Psychologica, 153, 95-106. DOI: 10.1016/j.actpsy.2014.08.001.

Abstract: A number of heuristic-based hypotheses have been proposed to explain how people solve syllogisms with automatic processes. In particular, the matching heuristic employs the congruency of the quantifiers in a syllogism – by matching the quantifier of the conclusion with those of the two premises. When the heuristic leads to an invalid conclusion, successful solving of these conflict problems requires the inhibition of automatic heuristic processing. Accordingly, if the automatic processing were based on processing the set of quantifiers, no semantic contents would be inhibited. The mental model theory, however, suggests that people reason using mental models, which always involves semantic processing. Therefore, whatever inhibition occurs in the processing implies the inhibition of the semantic contents.

We manipulated the validity of the syllogism and the congruency of the quantifier of its conclusion with those of the two premises according to the matching heuristic. A subsequent lexical decision task (LDT) with related words in the conclusion was used to test any inhibition of the semantic contents after each syllogistic evaluation trial. In the LDT, the facilitation effect of semantic priming diminished after correctly solved conflict syllogisms (match-invalid or mismatch-valid), but was intact after no-conflict syllogisms. The resultssuggest the involvement

of an inhibitory mechanism of semantic contents in syllogistic reasoning when there is a conflict between the output of the syntactic heuristic and actual validity. Our results do not support a uniquely syntactic process of syllogistic reasoning but fit with the predictions based on mental model theory.

Keywords: Inhibitory mechanism, matching heuristic, dual processing theories, mental model theory, syllogistic reasoning

7.1 Research background

Syllogisms are logical arguments comprising two premises and a conclusion. For example,

	Abstract	Concrete
Premise 1:	All A are B	All dogs are mammals
Premise 2 :	No B are C	No mammals are reptiles
----------------	----------------	----------------
Conclusion:	No A are C	No dogs are reptiles

There are 4 possible quantifiers per premise and conclusion:

Abbreviation	Quantifier	Example
A	universal affirmative	All A are B
E	universal negative	No A are B
I	particular affirmative	Some A are B
O	particular negative	Some A are not B

The conclusion is composed of two terms which we refer as “A” and “C”, and they appear in the first and second premises respectively. In both premises, there is a connecting term which we refer as “B”. The “B” term does not appear in the conclusion. The term ‘mood’ refers to the different combinations of quantifiers within the premises and conclusion. Therefore, the syllogism above has the mood AE-E. The 64 possible combinations of the mood together with the four possible figures (see Appendix E) yield a total of 256 syllogisms. However, among these 256 possible syllogisms, only 27 are valid (10.5%).

Syllogistic reasoning encapsulates many aspects of day-to-day reasoning, which involves the manipulation and transformation of our stored knowledge and information to make inferences about the world. In the above example, the knowledge concerned is our assumptions about category membership. However, even though this kind of reasoning is very common in day-to-day thinking, the average accuracy of syllogistic problems is only around 50% and the accuracy can be as low as 5-30% for the most difficult problems. Many studies have found evidence supporting the hypothesis that people do not often reason logically but rather they use some heuristic strategies, for example, the atmosphere (e.g., Begg & Denny, 1969; Woodworth & Sells, 1935) and the matching hypotheses (García-Madruga, 1983; Wetherick, 1989).

The atmosphere heuristic (Begg & Denny, 1969) leads to a conclusion that can be logical or not. There are two main assumptions in the atmosphere hypothesis:

1. Principle of quality: if there is at least one negative premise ('No' or 'Some...are not'), a negative conclusion is favoured; otherwise, a positive conclusion is preferred ('All' or 'Some').
2. Principle of quantity: if there is at least one particular premise ('Some' or 'Some...are not'), a particular conclusion is favoured; otherwise, a universal conclusion is preferred ('All' or 'No').

The matching hypothesis is a modified version of the atmosphere hypothesis which suggests that people choose the conclusion which matches the quantifier of the more ‘conservative’ premise, the premise with a lower number of entities. The suggested conclusion has the same quantifier (matches) as at least one of the premises, favouring the particular over the universal and negative rather than affirmative, i.e. E > O = I >>A (e.g., García-Madruga, 1983; Wetherick, 1989; Wetherick & Gilhooly, 1995). In the above AE-premises example, according to the surface structure of the premises, participants would tend to produce or accept E-conclusions. The atmosphere and matching hypotheses predict the same conclusions except for IE and EI-premises in which participants tend to produce/accept O-conclusions as suggested by the atmosphere hypothesis but E-conclusions by the matching hypothesis[14]. (Wetherick & Gilhooly, 1995) have found that 25 of their 71 participants apparently used matching to solve a syllogistic construction task of 40 problems. Only 16 participants used logical means to solve the problems. Stupple and Waterhouse (2009) have found evidence of the matching effect in conclusion evaluation tasks.

This heuristic strategy is purely syntactic in nature, in which the processing involves bind manipulation or "matching" of the quantifiers of the two premises and conclusion. Presumably, the semantic contents of the syllogism are not involved in such processing. However, these hypotheses are difficult to falsify as most of the valid conclusions do agree with the hypotheses, with the exception of five of the 27 valid syllogisms (AA-I4, AE-O2, AE-O4, EA-O1 and EA-O4, see Appendix F). The effect may be just an "unfortunate coincidence" of this fact (Johnson-Laird, 2006). We will explore this claim on the basis of mental model theory in this article.

[14] As the matching and the atmosphere hypotheses have the same prediction for our stimuli, we will just use "matching" in the rest of the article. We will just use "matching" in the rest of the article.

Dual-Process Theories of Reasoning

Since the article of Wason and Evans (1975), an increasing number of authors have proposed that there are 2 types of processing (systems) when people reason (Evans, 1984, 2003, 2008, 2010, 2011; Evans & Over, 1996; Evans & Stanovich, 2013a, 2013b; García-Madruga, 1983, 1989; Gigerenzer & Goldstein, 1996; Sloman, 1996; Stanovich, 2004; Stanovich & West, 2000). Type 1 processing (also known as System 1) refers to the unconscious, associative, intuitive and rapid processes which give outputs that may be prone to the bias of common sense, beliefs and previous experience. It is relatively undemanding of cognitive resources and independent of fluid intelligence. Responses from Type 1 processes are quick but errors are, sometimes, inevitable. The matching heuristic process is one of the Type 1 processes. Type 2 processes (also known as System 2) are thought to be conscious, analytical, rule-based, slow and more demanding of cognitive resources. They operate with effort and control and develop over time in humans. To solve complicated problems successfully, reasoners have to go beyond the superficial Type 1 output, discard it and engage in Type 2 processing[15] (through cognitive decoupling and mental stimulation).

Traditionally, Type 1 processing is thought to be context-based, while Type 2 processing is abstract and context-free. Several researchers have proposed a control system or mechanism for the shift from Type 1 to Type 2 processing (Evans, 2009; Stanovich et al., 2011) in possible conflict resolution. They have mainly proposed a tripartite structure, a System 3, which deactivates Type 1 processing (System 1). However, the mechanism of this shift is still under

---

[15] We adopted the default-interventionist structure (Evans, 2007) in this article, though there are other proposals of how the two systems/types of processing work together, such as the parallel-competitive architecture (Sloman, 1996, Smith & DeCoster, 2000).

debate (c.f. Thompson, 2009, 2010). In everyday life, people tend to accept the output of Type 1 processing and only activate Type 2 processing in some special situations, such as being explicitly instructed to reason logically, (Evans, 2006; Verschueren et al., 2005). Due to the limitations of cognitive resources and other factors, Type 2 processing sometimes still gives wrong responses (Evans & Stanovich, 2013b).

Evans and Stanovich (2013a, 2013b) have pointed out that the degree of involvement of working memory (WM) is one of the main distinctions between Type 1 and 2 processing. A significant (higher) correlation of WM measures can be regarded as an *indirect* proof of the use of analytical processes. Individual differences in WM capacities have been shown to be associated with syllogistic reasoning performance (Gilhooly et al., 1993). For example, studies employing a dual-task paradigm have consistently reported a role for the central executive and verbal WM in syllogistic reasoning (Capon et al., 2003; Gilhooly et al., 1999). We will elaborate on this point later.

One illustration of the interaction between Type 1 and Type 2 processing in everyday reasoning can be observed in the belief bias effect. This happens when common belief and logic are in conflict. More specifically, for syllogisms with concrete terms, in addition to the validity of the syllogism, both premises and the conclusion can agree or contradict with common beliefs. For example, “All apples are red” is “unbelievable” because there are green apples but “All apples are fruits” is “believable” because apple is a typical example of a fruit. In most of the belief bias studies, researchers manipulated the believability and the validity of the conclusion: a believable conclusion is not necessarily valid and vice versa.

As reasoning by automatic heuristic processes is more effective by demanding less cognitive resources (Evans, 2003, 2008; Sloman, 1996), it is natural that oftentimes people use common sense over logical reasoning as the preferred way of heuristics. As a result, they tend to commit the mistake of accepting invalid believable conclusions but rejecting valid unbelievable ones.

In the dual-processing theory framework, we may attribute the "superficial" response of participants to the automatic activation of the common belief (knowledge) by Type 1 processing. However, when Type 2 processing is activated, the problem is decoupled from its contents in the WM and the abstract/decontextualized representations are manipulated independently. Therefore, the output is free from belief bias effect. However, to produce the correct analytic response (with Type 2 processing), the inhibition/deactivation of the Type 1 output is essential, and the effect is more pronounced for invalid (but believable) than valid (but unbelievable) problems. The notion that inhibition plays an important role in analytical response is supported by some studies (e.g., Moutier, Plagne-Cayeux, Melot, & Houde, 2006). De Neys and Franssens (2009) have also linked belief bias effect to the failure to inhibit the irrelevant activated information in the long-term memory. Furthermore, it is suggested that an inhibitory mechanism that blocks the Type 1 heuristic process is required (e.g., De Neys, 2006; Evans, 2003).

De Neys and Franssens (2009) and then Steegen and De Neys (2012) have shown that when there was a conflict between the validity of a conclusion and its believability, inhibition of the semantic contents of the conclusion was induced only after correct syllogistic responses. They used a subsequent lexical decision task (LDT) to evaluate the inhibition of the semantic contents in reasoning. In a classic LDT, the recognition/reaction time of a word preceded by its semantically related word (the prime) is usually faster due to the semantic priming effect. The activation of the previous word helps in the activation of the target word due to automatic activation spreading through the semantic network (e.g., Collins & Loftus, 1975; De Groot, 1983; Meyer & Schvaneveldt, 1971; Perea & Rosa, 2002). In their experiments, participants were asked to evaluate the validity of four believable and four unbelievable (AA-A) syllogisms as follows:

	Believable	Unbelievable
Premise 1:	All flowers need water	All African countries are warm
Premise 2:	Roses need water	Spain is warm
Conclusion:	Roses are flowers (invalid)	Spain is an African country (invalid)
Condition:	conflict (invalid-believable)	no-conflict (invalid-unbelievable)

After each syllogism, related (e.g. rose, petal, garden, flower, plant and bush) and unrelated words to the two terms, A and C, in the conclusion were presented to the participants in a LDT, in which participants had to determine whether the letter string was a real word or not. They have found that the recognition time for the related words after correctly solved conflict syllogisms was slower than that after no-conflict syllogisms. However, such a result was not observed after incorrect syllogisms. Participants were cued by Type 1 processes to accept the conclusions of believable syllogisms but reject the conclusion of unbelievable syllogisms, as mentioned above. When the believability and validity of the syllogism are in conflict, the output of a Type 1 heuristic is to accept the invalid believable conclusion and reject the valid unbelievable conclusion. Inhibition of the heuristic processes, which relies on the believability of the problems, is necessary for successful reasoning as the believability cues the incorrect response for conflict syllogisms. They explained this impairment of the lexical access as the inhibition of the access of the associated knowledge, on which the heuristic is based. Therefore, the recognition time for the related word was delayed. De Ney and collaborators described this hindrance of lexical access in the memory as a “negative by-product of the belief discarding process”.

In the current study, the target Type 1 heuristic process to be inhibited is not based on the believability of the premises and/or conclusion but on the congruency of quantifiers in the premises with that in the conclusion, in accordance with the well-known heuristics in reasoning: the matching heuristic. All our syllogisms were believable. For the conflict syllogistic problems, the expected conclusion from the automatic process was invalid or the conclusion with incongruent quantifier was valid; while for the no-conflict problems, the quantifier of the valid conclusions agreed with the matching heuristic and that of the invalid conclusions did not. The aim of this study was to investigate whether the surface structure of the syllogisms, the matching effect, would affect reasoning performance and whether inhibition would be induced in the conflict resolution and lead to an impairment of lexical access.

Until now, we have assumed that the matching hypothesis is true, and the matching of quantifiers in a syllogism leads to an automatic process. According to De Neys and collaborators, the inhibition which caused the impairment of lexical access was induced by a conflict which is semantic in nature (believability). By contrast, it can be assumed that a conflict which is syntactic in nature would not induce such a consequence. That is, if the conflict is not induced by the believability of the meaning of the conclusion but by the congruency of the quantifiers and the validity, would there be impairment of semantic access? The dual-process theories do not predict such results according to the logic of De Neys and collaborators because the Type 1 processing to be inhibited, the matching heuristic, is syntactic in nature. There would not be any semantic “by-product” due to the discarding of such a syntactic Type 1 process. Only if the Type 1 process implied semantic processing would we expect impairment of lexical access. Indeed, the alternative prediction is suggested by the mental model theory, which explains results in matching effect studies in a different way, as will be explained later in this article.

In the experiment, participants were given syllogistic problems such as the following:

	Example 1	**Example 2**
	<u>no-conflict (single-model)</u>	<u>conflict (three-model)</u>
Premise 1:	No object with wheels is a boat	All flowers need water
Premise 2:	All vehicles have wheels	No computer is a flower

The key manipulation was that participants were given conclusions that differed in two properties: whether they were congruent or incongruent according to the matching heuristic and whether or not they were valid. ‘Conflict’ problems included the ‘match-invalid’ (the quantifier of the *invalid* conclusion “matches” with the quantifier of the “more conservative” premise) and mismatch-valid problems (the quantifier of the *valid* conclusion “mismatches” with the quantifiers of the two premises). For example, we constructed a mismatch-invalid no-conflict problem and a match-invalid conflict problem by providing “Some vehicles are not boats” and “No computer needs water” as the conclusions of Examples 1 and 2, respectively. Matching predicts a rejection of the conclusion of Example 1 which is invalid (correct reject). However, matching predicts an acceptance response of the conclusion of Example 2, which is invalid (false alarm). Following the paradigm of De Neys and his colleagues, after each syllogistic evaluation trial, participants were given a LDT of six related words (water, thirst, sea, computer, screen and keyboard), six unrelated words and 12 non-words.

Mental model theory

Contrary to the rule-based theories, which suggest that reasoning is syntactic in nature, the mental model theory suggests that reasoning involves semantic processing. Representation of the premises is not propositional in nature. The theory suggests that there are three stages during reasoning, namely, premise encoding, premise integration and conclusion validation. In the first stage, both the terms and their categorical relations of the premises are converted to some abstract tokens (abstract concepts like negation and quantifiers are also represented). The mental models are assumed to be constructed and manipulated in the WM. In order not to overload the WM, reasoners try to represent as little information as possible in the mental models and thus the models only represent what is true according to the premises, but not what is false (Johnson-Laird & Byrne, 2002; Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 2000), and some linguistic properties of the premises may not be represented. Reasoners form the initial mental model of the syllogism in the second stage, which is akin to the Type 1 process in dual-process theories and does not load on WM (Khemlani & Johnson-Laird, 2012). The third stage is the only deductive process in reasoning that involves a model-based search for counterexamples to the initial model and the construction of alternative model(s). If a counterexample is found, reasoners have to repeat the whole process until they find a model/a set of models without counterexamples. It is also assumed that a major component of the reasoning process is non-verbal. The problem is more difficult if more alternative models have to be constructed, as reasoners have to maintain and consider all of them to justify the validity of the conclusion. This process loads on cognitive resources, particularly the WM, and takes time (Bell & Johnson-Laird, 1998; Copeland & Radvansky, 2004). Therefore, it leads to errors and inefficiency in reasoning when the reasoners have limited WM capacity to represent and manipulate mental models (Johnson-Laird & Byrne, 1991) and/or limited time to search for the counterexamples. In the current study, we tested the WM capacities of participants to evaluate this claim. If participants solved the problems by constructing mental models (analytical process), higher correlations with the WM measures should be found for multiple-model than single-model problems (to be elaborated later).

As mentioned above, errors in reasoning are due to mistakes in the construction of mental models and insufficient search for counterexamples. More specifically, mental model theory does not agree with the view that some people use only heuristics for reasoning (cf. the computational implementation of mental model theory- mReasoner, Khemlani & Johnson-Laird, 2012), contrary to the matching hypothesis. People always construct mental models of the problem, at least the initial model. Many studies by Johnson-Laird and his colleagues have found results that agree with the predictions of the mental model theory (Johnson-Laird, 1983; Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991). Going back to the examples above:

	<u>no-conflict (single-model)</u>	<u>conflict (three-model)</u>
Premise 1:	No object with wheels is a boat	All flowers need water
Premise 2:	All vehicles have wheels	No computer is a flower

Reasoners construct the initial models of the two syllogisms “No object with wheels is a boat”, “All vehicles have wheels” (Example 1) and “All flowers need water”, “No computer is a flower” (Example 2) as follows:

Example 1:

Wheels Vehicles

Wheels Vehicles

Wheels

Boat

Boat

Example 2:

Model 1			Model 2			Model 3		
Flowers	Water		Flowers	Water		Flowers	Water	
Flowers	Water		Flowers	Water		Flowers	Water	
	Water			Water	Computer		Water	Computer
		Computer			Computer		Water	Computer
		Computer						

According to mental model theory, people always construct mental models when reason and mental models contain the semantic contents of the syllogisms, as in the examples above. By contrast, the use of a matching heuristic involves solely the surface structure of the statements. Such a hypothesis does not predict an impairment of lexical access or any effect on the semantic contents of the problem for solving conflict problems, as semantic processing is not involved. However, the mental model theory does predict such an effect, as semantic contents are also represented in mental models. As all our conflict problems are three-model syllogisms, construction of the two alternative models (and thus the finding of counterexamples) is necessary to solve the syllogism successfully. The initial mental model, which corresponds to the Type 1 output, leads to an incorrect response. The search for counterexamples and the “fleshing out” of mental models, which correspond to Type 2 processes, lead to inhibition of the incorrect initial model and also its semantic contents. This inhibition of the initial model (the output of Type 1 processing) will cause an impairment of lexical access if mental models are semantic in nature. In other words, if reasoning involved the construction and search for counterexamples of mental models, an effect on the semantic contents would be found; otherwise, if people used a heuristic that was purely syntactic in nature, such as matching, effects on semantic content would not be expected. We predicted a “slower” recognition time of the related words in the LDT as a side-effect of the inhibition of the initial mental model during the analytical processing (the “fleshing out” process, construction of alternative models and search for counterexamples) to solve conflict problems (three-model problems).

Between the two types of conflict problem, we expected a stronger inhibition effect on the match-invalid than the mismatch-valid problems. As reasoners generally exhibit the tendency to accept a conclusion rather than reject it, along with the fact that the heuristic output of matching also cues participants to accept a conclusion, it would be more difficult for them to reverse their response from acceptance to rejection. More inhibition and effort would be expected. However, it would be easier for participants to solve the mismatch-valid problems as the aforementioned tendency does not exist for the mismatch problems.

In a ‘no-conflict’ problem (see Example 1 above), for example the ‘mismatch-invalid’ problem, the quantifier of the “more conservative” premise does not “match” with that of the conclusion. Matching does not predict this conclusion, which is invalid. In addition to the agreement between the prediction by matching and the actual validity of the syllogism, more importantly, all our no-conflict problems are single-model syllogisms. The initial model gives the correct response. No analytical reasoning processes are required and thus no inhibition of the mental model is induced. We predicted no impairment of lexical access of the related words in the LDT and thus a faster recognition time of the related words than the unrelated words after this kind of problem. This is due to the facilitation effect of semantic priming of related words in the syllogistic task.

In summary, as all the eight conflict problems are three-model syllogisms, we predicted an inhibition of the initial mental model during the analytical reasoning processes, which in turn would cause an impairment of lexical access, as mental models are semantic in nature. In contrast, as all the no-conflict problems are single-model syllogisms, the initial model is sufficient for the correct responses and thus no inhibition would be induced in the reasoning process. No effect on lexical access would be found.

As mentioned above, the limiting factor of analytical processing is related to WM capacities. Some studies (but not all) have found correlation between WM capacity and performances in reasoning tasks (Capon et al., 2003; Gilhooly et al., 1993). The mental model theory also predicts the importance of WM in reasoning, as mental models are constructed and manipulated in the WM. Its capacity is a general limiting factor for the generation of new structural representations (alternative mental models) and limits the maximum amount of bindings to be kept simultaneously. All the eight conflict problems in this study were multiple-model problems while all the eight no-conflict problems were single-model syllogisms. We predicted that participants with higher WM capacities would perform better in the syllogistic task, in accordance with the mental model theory. As the conflict problems were extremely difficult, participants with insufficient WM and inhibition ability might not be able to solve them by the construction and counterexample search of mental models, if they employed analytical processes. They might solve the problems by other processes rather than inhibiting the incorrect initial model. (De Neys et al., 2005a, 2005b) have found differences in performances between high and low WM participants in conditional reasoning. In their 2005 study, participants with high WM used their WM resources to inhibit the retrieval of invalid disablers[16]. Studies have

[16] A disabler in conditional reasoning refers to a case in which the consequence is prevented despite the presence of the cause. For example, if you fall from the 15th floor, you will die. A disabler in this conditional could be a gravity-free environment, a cushion on the floor, etc.

shown that people with higher WM capacities have higher inhibitory control (e.g., Hasher et al., 2007; Lustig, Hasher, & Zacks, 2007). People with less WM capacity and thus inhibition control may not have enough resources to inhibit the irrelevant material and may employ different processes from people with high WM (e.g., Aslan & Bäuml, 2011; Braver, 2012; Braver, Gray, & Burgess, 2007). We predicted a stronger inhibition of the conflict problems and thus a larger impairment of lexical access in the LDT for high WM participants.

We used 4 additional tasks to measure the WM capacities and also the inhibition ability of participants. The Stroop task was used to measure the inhibition ability of participants. The main process in this study was the inhibition of the incorrect initial model(s). We expected a high correlation of this task with participants' performance in solving conflict problems and also the priming effect of the conflict problems. As inhibition is a main executive function, we used the semantic anaphora task (Elosúa, Carriedo, & García-Madruga, 2009) to measure both the verbal WM capacity and the executive function, the operation span task (adapted from Turner & Engle, 1989) for the executive function and the letter task for the visuo-spatial WM capacity and the executive function (Shah & Miyake, 1996).

In summary, we tested whether the matching effect was based on a heuristic process employing the paradigm of De Neys and Franssens (2009). No lexical access impairment would be predicted if the heuristic process were based on matching the quantifiers. However, the mental model theory would predict the opposite. In addition, we predicted a correlation of the WM and inhibitory control measures with both the percentage of correct responses of the syllogistic task and the negative priming effect of the LDT. We also predicted that participants with higher WM capacities would have a greater impairment of lexical access of the related words in the LDT after correct conflict problems, while participants with low WM capacities might not have enough cognitive resources to exert a sufficiently strong inhibition in the LDT, or they may employ other processes.

7.2 Method

Participants

107 students (mean age = 22.34 years, SD = 4.43; 89 females and 18 males) studying at the University of Granada participated in return for course credits. All participants were native Spanish speakers without any previous training in logic.

Material

Reasoning task. Participants were presented with 16 syllogisms. The combinations of premises which allow the formation of two conflict (match-invalid and mismatch-valid) and two no-conflict (match-valid and mismatch-invalid) conditions are the AE, EA and AA[17] problems (see Appendix F for details). We chose the AE and EA syllogisms because two of their valid conclusions are in favour of the matching heuristic (E single-model conclusions) and two are contrary to the matching heuristic (O three-model conclusions). The quantifiers of the two premises are not the same. This helped prevent participants from guessing our manipulation of the syllogisms.

Half of the syllogisms had AE-premises while the other half had EA-premises. Eight syllogisms had E-conclusions; 6 had O-conclusions and 2 had A-conclusions. The number of problems was counterbalanced in terms of validity, matching and conflict or not (2 x 2 x 2 design). Half of the syllogisms were conflict problems, in which four had the conclusion

[17] In the four AA syllogisms, two of them have a single-model valid conclusion (AA-A), one has no valid conclusion (AA-NVC) and the remaining one has a three-model conclusion (AA-I).

consistent with matching though they were in fact logically invalid (match-invalid problems) and the other four had the conclusion inconsistent with the matching though they were logically valid (mismatch-valid problems). The remaining eight syllogisms were no-conflict problems in which four had a logically valid conclusion consistent with the matching hypothesis (match-valid problems) and the other four had a logically invalid conclusion inconsistent with the matching hypothesis (mismatch-invalid problems). In other words, for the eight syllogisms with the consistent E-conclusion, four of them were conflict problems (invalid but consistent with matching) while the remaining four were no-conflict problems (valid and consistent with matching). It was the same for the inconsistent O-conclusion syllogisms. However, we replaced two of the invalid syllogisms with the A-conclusion, yielding four conflict O-conclusion syllogisms (valid but inconsistent), two no-conflict invalid O-conclusion syllogisms (i.e. mismatch-invalid) and two invalid A-conclusion syllogisms. The A-conclusion syllogisms were introduced as fillers (see Table 16). The properties of the A-conclusion fillers will be discussed in the Discussion section. All the premises and conclusions of the syllogisms were deliberately constructed to be believable in order to control the belief bias effect. The believability of the syllogisms was checked by 6 native speakers of Spanish. See Appendix G for the complete set of syllogisms.

Lexical decision task. In each trial of the lexical decision task, 12 Spanish words and 12 non-words were presented straight after the reasoning task of each syllogism (192 words and 192 non-words in total). Six of the Spanish words were words related to the two terms in the conclusion while the other six were words unrelated to the terms in the syllogisms. The words were selected according to the word frequency/number of production (frequent words were

Table 16

Types of Problem Used in the Experiment

Problem type	Conclusion type	Syllogisms
8 Conflict problems (multiple-model)	4 Match-invalid	2 AE-E(2) 2 EA-E(1)
	4 Mismatch-valid	2 AE-**O**(2) 2 EA-**O**(1)
8 No-conflict problems (single-model)	4 Match-valid	2 AE-**E**(1) 2 EA-**E**(2)
	4 Mismatch-invalid	AE-A(1)* AE-O(1) EA-O(2) EA-A(2)*

* Filler problems

Note. Figure in parenthesis. Valid conclusions are in bold.

selected) and the direct association indices from the NIPE Spanish words database created by the University of Salamanca (http://campus.usal.es/~gimc/nipe/). Six native speakers of Spanish were asked to check all the stimuli. Twenty-three participants were asked to complete a survey on the association index of the words to the two terms in the conclusions and the subjective word frequency on a 10-point scale. The overall word frequency was 6.37 (related = 6.30; unrelated = 6.43). The association index of the related words (except the two terms in the conclusion) was 8.23 while that for the unrelated words was 0.46. See Appendix G for a complete list of the related and unrelated words.

Procedure

1. Syllogistic conclusion evaluation task and the lexical decision task. Participants were tested individually in a quiet room sitting in front of two computers. In each trial, they had to evaluate whether the given conclusion of the syllogism was valid or not. Right after the syllogistic evaluation task, they had to decide whether the 24 consecutively presented strings were words or not. Two practice trials were presented before the 16 experimental trials. The standard deductive reasoning instruction was used. Participants were asked to assume that the two premises were always true and the conclusion was valid only if it followed logically from the two premises. An example of a valid AA-A (of Figure 1) problem with unbelievable conclusion was used as an illustration. In the syllogistic reasoning task, each premise was presented consecutively for 3 seconds followed by 500ms of a blank page. The two premises and the conclusion were then displayed until the participant pressed a key to indicate his/her response. There was no time limit for the response. Participants were asked to respond as accurately as possible. The first letter string was then displayed after a blank page of one second. The letter string was replaced by the next letter string once the participant responded. The maximum response time was 5 seconds for each letter string in the LDT. The 16 problems and the corresponding letter strings were presented in random order. Participants were asked to respond as fast and as accurately as possible for the LDT.

2. Working Memory and executive processes tasks. After the above tasks, participants were asked to complete three WM tasks and the Stroop task. The order of presentation was counterbalanced among participants. The three WM tasks included the Letter Task, Semantic Anaphora Task and Operation Span Task.

Letter Task. This task was adapted from the one in Shah and Miyake (1996) which aimed to test the visuospatial WM of the participant, with a component of the central executive. Each item comprised one of five letters (F, J, L, P, or R), either as a normal or mirror image of the letter (i.e. R and Я respectively), presented in one of seven possible orientations, except the upright position (i.e. 45º, 90º, 135º, 180º, 225º, 270º and 315º). Test items were presented on a computer in sets of ascending size, from two to five items, with five sets at each level, 20 sets in total. The presentation was constrained so that opposing orientations (e.g. 45º and 225º) were not presented successively within a set and each orientation appeared only once per set. Each letter remained on the screen for a maximum of 2200 ms or until the participant responded. Participants had to press "P" if the figure was a normal image of the letter and press "Q" if it was a mirror image. After each set, participants had to indicate the orientation of the top of each letter, in the order the letters were presented. One point was awarded per letter for each correct set.

Semantic Anaphora task. This task (in Spanish) was developed by {Elosúa, Carriedo and García-Madruga (2009). It tests the (verbal) WM of participants. The dual task design aims to measure the executive control capacity of the WM. Firstly, participants have to solve the pronominal anaphora problems and then remember the word-solution of the problems. For example:

Eladio encouraged **her** a lot to interpret such a demanding role

[Eladio **la** animó mucho a que interpretara aquel papel tan exigente]

career   actress

[carrera] [actriz]

In this case, both words, “career” and “actress”, are grammatically appropriate, because the pronoun “**la**” (“her” in English) is used in Spanish to refer to both feminine terms and people. However, only the word “actress” is the correct response because it matches the meaning of the sentence.

The participant had to read each sentence aloud and remember the word-solutions of the anaphora. Sentences were presented on a computer in sets of increasing size from two to five sentences with 3 sets in each level (12 sets in total). Participants were asked to recall the words in the correct order after each set of sentences. The maximum span was recorded.

Operation span task. This task (adapted from Turner & Engle, 1989) is strongly related to the central executive capacity. In each trial, participants had to verify a mathematical operation and then memorise a monosyllabic word (in Spanish).Three sets from levels two to six were presented in random order (15 sets in total). Participants had to report all the words in the correct order after each set. One point was awarded per word for each correct set.

Stroop Task. Participants were asked to read aloud the colour of the words. There were 75 trials in total, with 25 consistent trials (the colour of the word was the same as the meaning of the word), 25 inconsistent trials (the colour of the word was different from the meaning of the word, for example ‘purple’ in yellow) and 25 control trials (non-colour words, for example ‘weight’, in green). This task aims to measure the inhibition ability of the central executive function and the participants’ ability to resist the interference.

7.3 Results

Syllogistic Evolution Task

The conflict problems generally had a lower accuracy. Excluding the filler A-conclusion problems, conflict problems had a comparatively lower accuracy, 27%, than no-conflict problems, 85%, t(106) = -26.149, $p < .001$.

Match problems had an accuracy of 51% while mismatch problems had an accuracy of 53%. Valid problems had an accuracy of 70% while invalid problems had an accuracy of 28%, t(106) = -16.987, $p < .001$. The 2 (type of conclusion: match vs. mismatch) X 2 (validity: valid vs. invalid) ANOVA showed a significant main effect of type of conclusion (match vs. mismatch), F (1, 106) = 9.965, $p = .002$, $\eta_p^2 = .086$; and validity, F (1,106) = 138.732, $p < .001$, $\eta_p^2 = .567$; and a significant interaction effect of the above two factors, F (1, 106) = 443.397, $p < .001$, $\eta_p^2 = .807$. See Table 17 for the details.

Participants had a longer response latency (RT) for conflict problems than no-conflict problems, 17526ms and 7507ms, respectively, when incorrect responses were excluded, t (99) = 7.676, $p < .001$.When incorrect responses were *not* excluded, the RT of the syllogistic task for conflict and no-conflict problems were 11939ms and 8182ms, respectively, t(106) = 8.604, $p < .001$. More specifically, participants had a shorter RT for match-valid problems (6383ms)[18] than match-invalid problems (10926ms) and mismatch-invalid problems (11576ms) than mismatch-valid problems (18388ms), t (20) = 2.149, $p = .04$ and t (85) = 3.509, $p < .001$.

[18] The values are different from Table 17 because participants who did not have at least one correct response to each of the problems were excluded in the t-tests.

Table 17

Percentages and Latencies of Correct Responses in the Syllogistic Task

Problem type	Conclusion type	Accuracy	RT (ms)
8 Conflict (total)		27%	17526 (15277)*
	4 Match-invalid (AE/EA-E)	7%	10926 (8691)
	4 Mismatch-valid (AE/EA-**O**)	46%	18017 (16509)
6 No-conflict (total)		85%	7507 (4180)
	4 Match-valid (AE/EA-**E**)	94%	6047 (4064)
	2 Mismatch-invalid(AE/EA-O)*	68%	11519 (7875)

* standard deviation in parentheses.

Note. Valid conclusions are in bold.

Correlation analysis. The facilitation index of the Stroop task was calculated by subtracting the reaction time of the congruent items from the control items (non-colour words), while the inhibition index of the Stroop task was calculated by subtracting the reaction time of the control items from incongruent items. The sum of the z-scores of letter score, operation span score and semantic anaphora span was calculated and used in the correlation analysis as the WM composite. For the correlation analysis of the reasoning task with the WM and executive function measures, the overall accuracy in the syllogistic reasoning task correlated with most of the WM measures (letter score: r (99) = .281, p = .005; operation score: r (102) = .235, p = .017, and semantic anaphora span, r (102) = .208, p = .036). In particular, the correlation between overall accuracy and WM composite was r (102) = .350, $p < .001$. However, the overall accuracy did not correlate with the Stroop task measures (see Table 18).

Table 18

Correlations Between the Accuracies of the Syllogisms and the Working Memory Measures

		Stroop_ Facil.	Stroop_ Inhibition	Letter Score	Operation Score	Anaphora Span	WM composite
Conflict		**.203****	-.157	.075	.112	.143	.175#
	Match-invalid	.066	-.004	.156	.142	.099	**.218***
	Mismatch-valid	.189#	-.179#	-.006	.033	.097	.053
No-conflict		-.047	.193#	**.315****	**.215***	.133	**.309****
	Match-valid	-.043	.124	**.218***	.133	-.026	.153
	Mismatch-invalid	-.025	.148	**.225***	.169#	**.198***	**.275***
overall excl. A-concl.		.143	-.009	**.281****	**.235***	**.208***	**.350****

** p < .01; * p < .05; # p < .10; two tailed. incl. = including, excl. = excluding, A-concl = A-conclusion problems.

The accuracy of the conflict problems correlated with the facilitation effect of the Stroop task, r (102) = .203, $p = .041$ and there was also a marginally significant correlation with the WM composite, r (102) = .175, $p = .078$. When dividing the conflict problems into the two specific problem types, only the match-invalid problems correlated with the composite measure of WM, r (102) = .218, $p = .028$; the mismatch-valid problems did not correlate significantly with any of the measures. The accuracy of the no-conflict problems correlated with the letter score, r (99) = .315, p = .002, the operation score, r (102) = 215, p = .030 and the WM composite, r (102) = .309, p = .002. See Table 18 for the correlations. Dividing the no-conflict problems, the match-valid problem correlated with the letter score, r (99) = .218, p = .030; and the mismatch-invalid problems correlated with the letter score, r (99) = .225, p = .025, semantic anaphora span, r (102) = .198, p = .046, and the WM composite, r (102) = .275, p = .005.

Lexical Decision Task

Participants obtained a very high accuracy for the LDT. The accuracies for non-words, related words and unrelated words were: 95%, *SD* = .212; 98%, *SD* = .135; and 97%, *SD* = .172, respectively. Incorrect responses in the LDT were excluded in the following analyses, as in De Neys and Franssens (2009).

After correct syllogisms[19], related words of the conflict problems had a similar reaction time in the LDT to the unrelated words, 695ms and 693ms, respectively. On the other hand, after correct no-conflict syllogisms, related words had a faster reaction time than unrelated words, 692ms and 730ms, respectively, t(106) = 5.555, $p < .001$. The results suggested that the semantic priming effect was eliminated after solving the conflict syllogisms correctly. In the 2 (conflict vs. no-conflict) X 2 (related vs. unrelated word) ANOVA, the main effect of conflict status was not significant but the main effect of word type and the interaction effect of conflict status and word type were significant, F (1, 98) = 5.316, p = .023, η_p^2 = .051 and F (1, 98) = 7.049, p = .009, η_p^2 = .067, respectively. The significant main effect of word type was as expected, as related words were primed semantically by the previously presented syllogisms. One interesting observation was that the reaction time of the unrelated words of the conflict problems was *faster* than that of

[19] We use the term "correct syllogism" to refer to those syllogisms which participants gave a correct response in the syllogistic task.

the no-conflict problems, t (98) = 3.078, $p = .003$.We will try to explain this in the Discussion section.

De Neys and Franssens (2009) have found significant effects for the two main factors and the interaction effect. We therefore carried out an analysis classifying the conflict problems into match-invalid and mismatch-valid and the no-conflict problems into match-valid and mismatch-invalid problems. This was meant to check whether the non-significance of the main effect of the problem type (conflict vs. no-conflict) was due to the cancellation of the effects by the two sub-types of the conflict problems and no-conflict problems. Only the interaction effect of the three factors, conclusion type (match vs. mismatch) x validity (valid vs. invalid) x word type in the LDT (related vs. unrelated) was significant in the 2x2x2 ANOVA, F(1, 16) = 7.113, $p = .017$, $\eta_p^2 = .308$. The t-tests showed that after correct syllogisms, for the two no-conflict conditions, recognition time of the unrelated words was significantly slower than that of the related words (match-valid: t(106) = 5.089, $p < .001$; mismatch-invalid: t (92) = 2.643, $p = .010$). Match-invalid was the only condition where the lexical decision time of the related words was slower than that of the unrelated words, though it was marginally significant, t (19) = 1.850, $p = .080$. The small *p*-value may be due to the small *n* after excluding incorrect syllogistic trials. For mismatch-valid problems, the reaction times of the related and unrelated words were comparable. See Table 19 for a summary.

However, after **incorrect** syllogisms, the reaction time of the unrelated words was generally slower than that of the related words even after answering conflict problems, t(106) = 2.335, $p = .021$.

Correlation analysis of the lexical decision task. The semantic priming effect of the related words was calculated by subtracting the reaction time of the unrelated words from that of

Table 19

Summary of the Lexical Decision Times

		Correct syllogism		Incorrect syllogism	
		Related	Unrelated	Related	Unrelated
Conflict		**695**	**693**	**681**	**695***
	Match-invalid	815	725#	677	692*
	Mismatch-valid	689	688	692	704#
No-conflict		**692**	**730****	**685**	**701**
	Match-valid	692	730**	739	734
	Mismatch-invalid	689	722**	676	700
Total general		***695***	***718****	***684***	***698****

** p< .01; * p < .05; # p< .10; two tailed. Please note that there were unequal numbers of participants. In each cell participants had a different number of correct and incorrect trials. Steegen and De Neys (2012) applied the same classification (correct and incorrect syllogistic responses). The lack of significance of some of the t-tests may be due to the small number of *n* after eliminating the correct or incorrect trials.

the related words. The value should be positive for a facilitation effect, i.e. faster recognition time of the related words. The priming effect of the conflict problems in the LDT correlated with both the facilitation index and the inhibition index of the Stroop task, r (94) = -.253, p = .014 and r (94) = .204, p = .048, respectively. For the WM measures, the priming effect of conflict problems correlated with the score of the operation span and WM composite, r (94) = -.250, p = .015 and r (94) = -.210, p = .042, respectively. The priming effect of no-conflict problems did not correlate significantly with any of the measures. Dividing the conflict problems into match-invalid and mismatch-valid problems, only the mismatch-valid problem correlated significantly with both the Stroop facilitation and inhibition, r (93) = -.233, p = .025; r (93) = .231, p = .026, respectively. The mismatch-valid problem did not correlate significantly with any of the measures, probably due to the fact that there were only 18 participants who had non-zero accuracy for this problem type. See Table 20 for a summary.

Working Memory Capacities

According to the mental model theory, having a high WM capacity is important for solving multiple-model problems with analytical processes. We expected different reasoning processes for participants with different WM capacities. We selected the 30% of participants with highest and lowest WM composite scores, 34 participants in each group, for the following t-tests. Participants with high WM, surprisingly, did not perform significantly better than those with low WM for conflict problems (high: 28%; low: 21%, t (66) = 1.782, p = .079) but they did so for no-conflict problems (high: 90%; low: 81%, t (65) = 2.501, p = .015). More importantly, they had different results for the LDT, see Figure 1. The results showed that participants with high WM had a different priming effect (difference between the recognition times for related words and unrelated words) after correct conflict than no-conflict syllogisms (priming conflict: -26ms; priming no-conflict: 30ms, t (31) = 3.004, p = .005) but participants with low WM did not (priming conflict: 26ms; priming no-conflict: 22ms, t (28) = 0.137, p = .892).

Table 20

Correlation Analysis of the Lexical Decision Task

		Stroop_ Facil.	Stroop_ Inhibition	Letter Score	Operation Score	Anaphora Span	WM composite
Priming							
conflict		**-.253***	**.204***	-.086	**-.250***	-.085	**-.210***
	match-invalid	-.211	-.050	-.077	-.276	.198	-.137
	mismatch-valid	**-.233***	**.231***	-.096	-.170	-.090	-.163
Priming							
no-conflict		-.163	-.001	.045	-.008	-.048	-.018
	match-valid	-.099	-.043	.016	-.044	-.062	-.056
	mismatch-invalid	-.158	.062	.130	.071	.034	.109

* $p < .05$, # $p < .10$; two tailed

Note. Only participants with at least one correct syllogistic response were included in the analysis and thus there were only 18 participants in the priming effect of match-invalid problems. The Pearson correlations did not reach significant level though two of the values were fairly high.

7.4 Discussions

The main aim of this research was to check whether the cause of the inhibition of semantic contents in syllogistic reasoning during conflict resolution could be extended to the congruency of the quantifiers, which may be thought of as part of the surface structure. We have found a diminished/negative priming effect on the semantic contents of syllogisms in correct conflict syllogisms. More specifically, we found a negative semantic priming effect for match-invalid problems and a null result (i.e., the absence of priming) for mismatch-valid problems. In line with the findings of previous studies by {De Neys and his colleagues (2009, 2012}, after correct conflict syllogisms, the semantic priming facilitation of the related words diminished, especially for the match-invalid condition. The facilitation effect of semantic priming in the mismatch-valid problems was cancelled out. Moreover, the recognition time of the related words after correct conflict invalid syllogisms with conclusions congruent with the matching hypothesis (match-invalid) was (marginally) significantly slower than that of unrelated words ($p = .080$). De Neys and collaborators explained the impairment of lexical access as a “by-product of the Type 1 processing discarding process”, according to the dual-process theories. However, the heuristic process in our study was syntactic in nature and therefore, contrary to our results, no lexical access impairment would be expected: inhibition of a process which was not semantic in nature did not logically lead to an inhibition of the semantic contents.

The result was against a simple matching heuristic process based on the processing of quantifiers (all, some, some not and none) and the view based on propositional reasoning. The induced inhibitory mechanism did not affect only the process of matching quantifiers (a syntactic process) but also the terms (e.g. “flowers”) in the syllogisms. The mental model theory gives an alternative explanation to the traditional results obtained with the matching heuristic, asserting that the difficulty of conflict problems is not based on the mismatching of quantifiers but because they require multiple mental models (our conflict problems were multiple-model syllogisms while our no-conflict problems were single model syllogisms). The present results fit with the view that there was a semantic priming effect for single-model (no-conflict) syllogisms but not for multiple-model (conflict) syllogisms. For multiple-model problems, the analytical reasoning process involved the discarding of the automatic initial model, the construction of alternative model(s) and the search for counterexamples. Inhibition of the initial model and its semantic contents was induced during the processing. An impairment of the lexical access of the two terms in the conclusions was induced after the successful solving of conflict problems, as a by-product of the inhibition of the initial mental model, even when the conflict was apparently induced by the surface structure of the syllogism. Decision time in the syllogistic task was shorter for match-valid than match-invalid problems and mismatch-invalid than mismatch-valid problems, again supporting our hypothesis that the construction of additional mental models was necessary to solve our multiple-model problem, as more time was needed for the processes.

For the three-model match-invalid problems, inhibition of the initial model during analytical processes was expected to successfully solve the problem. The inhibition of the initial mental model as well as lexical access to the words (as the by-product) caused the retrieval of the related words to be more time-consuming than that of unrelated words in the same condition and related words in the no-conflict condition (815ms for related words in the match-invalid problems vs. 725ms for unrelated words in the match-invalid problems vs. 692ms for related words in the match-valid problems). More effort and thus time was required to activate the previously inhibited words in the subsequent LDT.

As mentioned before, reasoners generally have the tendency to accept a conclusion (Evans et al., 1983; Morley et al., 2004), disregarding the actual validity. In our results, 70% of the valid syllogisms were correctly solved but the percentage of correct responses for invalid problems was only 42%. For valid conclusions, participants could give the correct response by blind acceptance of the (valid) conclusion without any analytical processes. The match-invalid problems were more difficult than the mismatch-valid problems as participants could not give the correct response by blind acceptance of the conclusion. The accuracies of these two kinds of problems were 10% and 42%, respectively, underlining our expectation that match-invalid problems were more difficult. It was possible that participants employed neither the matching nor constructing mental models for a correct response to mismatch-valid problems. If participants did not employ analytical processes for solving the problems, no/little inhibition was induced. The results of this condition were thus more controversial and more prone to individual differences.

For incorrect syllogisms, it could not be ensured that participants actually processed the contents of the problem. In fact, we found no significant difference between related and unrelated words; however, it was significant after correct no-conflict syllogisms. When reasoners did not solve the syllogisms correctly, deep processing could not be guaranteed. If they were not engaged in solving the problems, even if there were some pre-activation of the related words, the activation levels would be less than those after correct syllogistic responses. In other words, the differences in the activation levels of the related and unrelated words could be less than that after correct no-conflict syllogisms.

Previous studies have demonstrated the relationship between WM and reasoning (Copeland & Radvansky, 2004; De Neys et al., 2005b). Copeland and Radvansky (2004) have shown that people with larger WM spans perform better in syllogistic reasoning. As predicted, we found a reliable correlation pattern between overall percentage of correct responses of the syllogistic task and WM measures.

It has been found that more cognitive resources are required to solve conflict problems for processes such as inhibition of the salient heuristic Type 1 response (De Neys & Van Gelder, 2009; Handley et al., 2004; Houdé, 1997; Moutier et al., 2006; Simoneau & Markovits, 2003). The performance of participants in solving conflict problems in the syllogistic task correlated with the Stroop task. Participants had to inhibit the unrelated materials as well as the irrelevant processes (such as the heuristic process) for successful reasoning. This view was further supported by the results of the LDT, as we will explain later. However, surprisingly, we found low and in most cases unreliable correlations between WM measures and participants' performances in solving conflict problems (the multiple-model problems which required analytical processing), whereas we found several significant correlations of different WM measures with no-conflict problems (see Table 18). A possible explanation was that the conflict problems were too difficult and solving them required WM capacities beyond the limitations of **some** participants; an increase in WM capacities did not improve the performance of solving these multiple-model problems and thus we found only a marginally significant correlation with the WM composite.

However, when we divided the conflict problems into their two types, the analysis of correlations between the percentage of correct responses and WM and Stroop measures in both kinds of conflict problems showed that most of the correlations were in the predicted direction/polarity (all were positive correlations except for the Stroop inhibition, though not significant) but the patterns for match-invalid and mismatch-valid problems were different. For the match-invalid (conflict) problems, there was a light but consistent positive correlational pattern with WM measures that facilitated a significant correlation with the WM composite, whereas the correlations with the two Stroop measures were non-significant. For the mismatch-valid (conflict) problems, there was a marginally significant pattern of correlations with the two Stroop measures, whereas the correlations with WM measures were close to zero. This differential pattern was unsurprising since the reasoning processing in these two types of problems was clearly different. In the match-invalid problems (AE/EA-E), reasoners had to carry out a deep analytical reasoning process in order to discover that the given E-conclusion was invalid. As the automatic-heuristic E-conclusion was presented to them explicitly to evaluate (a bottom up process), their task consisted of checking whether the conclusion was invalid and searching for a new valid output (alternative models). On the other hand, for mismatch-valid problems (AE/EA-O), reasoners had to inhibit the invalid match E-conclusion, the conclusion generated by their automatic-heuristic top-down process, in order to confirm that the given O-conclusion was valid. As shown in the accuracy results, participants' analytical reasoning process in mismatch-valid problems was easier than that in match-invalid problems because they just had to confirm that the given response was correct for the former problems.

Moreover, as we mentioned before, there was a tendency to accept or confirm conclusions, that is, some participants might blindly accept the conclusion without any analytical processes. This hypothesis was supported by the results of the LDT. We found a diminished priming effect for mismatch-valid problems but a negative result for the match-invalid problems. If the participants did not solve the problem with analytical processes, no inhibition and thus no impairment of the lexical access would be induced. The effect might be "neutralised" by those non-analytical participants so that the priming effect after correct mismatch-valid problems was non-negative. The absence of significant correlations for the mismatch-valid problems may be due to individual differences in reasoning processes.

In the no-conflict problems, we found a reliable correlation between accuracy in the reasoning task and most of the WM measures. Reasoning in single-model problems also required WM resources to convert the premises into mental models and construct the initial models. Although the initial model corresponded to the Type 1 output, in contrast to the assumption of the dual-process theories that Type 1 processing is relatively undemanding of cognitive resources, our results suggested the importance of WM for “automatic” processes in reasoning. The analysis of correlations between accuracy and WM measures in both kinds of no-conflict problems showed that the correlations were higher and more consistent in the mismatch-invalid than in the match-valid problems. As with the two types of conflict problem, reasoning processes in the two no-conflict problems were also different.

In match-valid (AE/EA-E) problems, the reasoners’ task was very easy; they only had to confirm that the given E-conclusion, which agreed with the matching heuristic, was valid. Only the complex visuo-spatial task correlated with the accuracy of this problem type and this correlation suggested that participants were building the correct mental model of the problem (consistent with the mental model theory’s proposal about the spatial (and iconic) nature of mental models). In other words, the correct valid response was probably yielded as a result of a semantic process, not just from the superficial heuristic matching process.

In the mismatch-invalid (AE/EA-O) problems, the reasoners’ task was somehow less easy. Participants had to conclude that the given response (O) was wrong, though with the help of the superficial matching heuristic, which cued them that the E-conclusion was valid. Reasoners had to construct the model of the correct response and decide that there was no other possible model. As the results have shown, WM resources seemed to be more involved in mismatch-invalid than in match-valid no-conflict problems.

The priming effect of the conflict problems in the LDT correlated with the two measures of the Stroop task, the operation span task and the WM composite. However, the priming effect of the no-conflict problems did not correlate with any of the WM and Stroop measures. The results point to the importance of the executive functions in solving conflict problems, which is in agreement with the common finding that executive functions are very important in reasoning, e.g. updating the representations, connecting with the long-term memory and inhibiting the automatic responses (García-Madruga et al., 2007). The correlations with the Stroop task measures indirectly supported our hypothesis that the diminished/negative semantic priming effect after correct conflict problems was the consequence of an inhibitory process during reasoning.

Our last hypothesis on the importance of individual difference in reasoning was that participants with low WM capacities (and thus low executive functions capacities) might not have sufficient cognitive resources for the demanding analytical processes needed to solve multiple-model problems. Participants with high WM capacities would be more likely to employ analytical reasoning processes. Our results showed that lexical access of the related words after correct conflict problems was impaired only for high WM participants. The results suggested that these two types of participants employed different reasoning processes to solve multiple-model syllogisms. Participants with high WM capacities had higher inhibitory control and thus a greater impairment of lexical access in the subsequent LDT.

Regarding the invalid A-conclusion fillers, we included them to boost the use of the matching heuristic. The conclusion was not consistent with matching. They should be regarded as no-conflict problems. Participants could reject the conclusion easily by noticing the incongruence of the “surface structure” of the premises and the conclusion. It was nonsense for the conclusion to be universal affirmative (A) as one of the premises was universal negative (E). This type of problem encouraged the use of the matching heuristic, forcing an analysis of the quantifiers. The high accuracy, 85%, of the A-conclusion problems suggested that participants had no problem rejecting the invalid conclusion (using the matching). However, the related words were recognised significantly more slowly than the unrelated words in the LDT (related words: 696ms, unrelated words: 662ms; $t(100) = 3.310$, $p = .001$). This result suggested that an inhibition mechanism was employed during the syllogistic task. It was natural and easy for any person to reject the conclusion if one of the premises states that ‘All A are B’ and the other states that ‘No B are C’; nobody would accept the conclusion ‘All A are C’ by instinct. The tricky point of the problem was that one of the premise statements was an atypical exemplar of the category. The two A-conclusions were as follows:

All bats are mammals
<u>No mammal can fly</u>
All bats can fly

No plant eats insects
<u>All carnivorous plants eat insects</u>
All carnivorous plants are vegetables

For both syllogisms, participants had to inhibit the semantic of the first premise in order to accept/assume that the second premise was always true (participants were instructed to assume that the two premises were always true). For the first syllogism, bats are an atypical exemplar of mammals which can fly. When participants read the second premise ‘no mammal can fly’, they had to inhibit the previous premise that ‘bats are mammals’, as bats can fly. Similarly, participants had to inhibit their knowledge that ‘no plants eat insects’ before making the assumption that ‘all carnivorous plants eat insects’ was always true. This inhibition then affected the performance of participants in the LDT.

In conclusion, the current study contributes to the literature by indicating that the matching heuristic in syllogistic reasoning is not only a syntactic-propositional effect based on quantifiers. A conflict induced by the congruency of the quantifiers in syllogisms would indeed cause an impairment of lexical access of the semantic contents. The effect can be explained, according to the mental model theory, by the number of mental models of the syllogisms. There is an undeniable involvement of semantic processing in reasoning. In this case, the impairment of lexical access is a by-product of the inhibition of the initial mental model during the analytical reasoning process. However, the actual locus of this inhibitory mechanism remains unknown. Possible further studies could be about the inhibition mechanism of the AE-A syllogism with an atypical exemplar as one of the premises and also the difference between matching and atmosphere heuristics.

Reference

Aslan, A., & Bäuml, K.-H. T. (2011). Individual differences in working memory capacity predict retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 264.

Begg, I., & Denny, J. P. (1969). Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning errors. *Journal of Experimental Psychology: , Memory, and Cognition, 81*(2), 351-354.

Bell, V. A., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science, 22*(1), 25-51. doi:10.1016/S0364-0213(99)80034-2

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences, 16*(2), 106-113.

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. *Variation in working memory,* 76-106.

Capon, A., Handley, S., & Dennis, I. (2003). Working memory and reasoning: An individual differences perspective. *Thinking & Reasoning, 9*(3), 203-244.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407.

Copeland, D., & Radvansky, G. (2004). Working memory and syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 57*(8), 1437-1457. doi:10.1080/02724980343000846

De Groot, A. M. (1983). The range of automatic spreading activation in word priming. *Journal of Verbal Learning and Verbal Behavior, 22*(4), 417-436.

De Neys, W. (2006). Automatic–heuristic and executive–analytic processing during reasoning: Chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology, 59*(6), 1070-1100.

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition, 113*(1), 45-61.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005a). Working memory and counterexample retrieval for causal conditionals. *Thinking & Reasoning, 11*(2), 123-150.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005b). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning, 11*(4), 349-381.

De Neys, W., & Van Gelder, E. (2009). Logic and Belief across the Lifespan: The Rise and Fall of Belief Inhibition during Syllogistic Reasoning. *Developmental Science, 12*(1), 123-130. doi:10.1111/j.1467-7687.2008.00746.x

Elosúa, M. R., Carriedo, N., & García-Madruga, J. A. (2009). Dos nuevas pruebas de Memoria Operativa de Anáforas. *Infancia y aprendizaje, 32*(1), 97-118.

Evans, J. S. B. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75*(4), 451-468.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10).

Evans, J. S. B. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*(3), 378-395.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol., 59*, 255-278.

Evans, J. S. B. (2009). How many dual-process theories do we need? One, two, or many?

Evans, J. S. B. (2010). Intuition and reasoning: A dual-process perspective. *Psychological Inquiry, 21*(4), 313-326.

Evans, J. S. B. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review, 31*(2), 86-102.

Evans, J. S. B., Barston, J., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition, 11*(3), 295-306. doi:10.3758/BF03196976

Evans, J. S. B., & Over, D. E. (1996). Rationality in the selection task: Epistemic utility versus uncertainty reduction.

Evans, J. S. B., & Stanovich, K. E. (2013a). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science, 8*(3), 223-241.

Evans, J. S. B., & Stanovich, K. E. (2013b). Theory and metatheory in the study of dual processing: Reply to comments. *Perspectives on psychological science, 8*(3), 263-271.

García-Madruga, J. A. (1983). Un modelo general sobre el razonamiento silogístico: Doble procesamiento y fase de comprobación con verificación. *Revista de Psicología General y Aplicada*.

García-Madruga, J. A. (1989). Comprensión e inferencia en el razonamiento silogístico. *Cognitiva, 2*(3), 323-350.

García-Madruga, J. A., Gutiérrez, F., Carriedo, N., Luzón, J. M., & Vila, J. O. (2007). Mental models in propositional reasoning and working memory's central executive. *Thinking & Reasoning, 13*(4), 370-393.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review, 103*(4), 650.

Gilhooly, K. J., Logie, R., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & Cognition, 21*(1), 115-124.

Gilhooly, K. J., Logie, R., & Wynn, V. (1999). Syllogistic reasoning tasks, working memory, and skill. *European Journal of Cognitive Psychology, 11*(4), 473-498.

Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. S. B. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning, 10*(2), 175-195.

Hasher, L., Lustig, C., Zacks, R., Conway, A., Jarrold, C., & Kane, M. (2007). Inhibitory mechanisms and the control of attention. *Variation in working memory, 19*, 227-249.

Houdé, O. (1997). The problem of deductive competence and the inhibitory control of cognition.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*: Harvard University Press.

Johnson-Laird, P. N. (2006). *How we reason*. Oxford: University Press.

Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition, 16*(1), 1-61.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*: Lawrence Erlbaum Associates, Inc.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review, 109*(4), 646.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., & Legrenzi, M. S. (2000). Illusions in reasoning about consistency. *Science, 288*(5465), 531-532.

Khemlani, S., & Johnson-Laird, P. N. (2012). Theories of the Syllogism: A Meta-Analysis. *Psychological Bulletin, 138*(3), 427-457. doi:10.1037/a0026841

Lustig, C., Hasher, L., & Zacks, R. T. (2007). Inhibitory deficit theory: Recent developments in a “new view”. *Inhibition in cognition, 17*, 145-162.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227.

Morley, N. J., Evans, J. S. B., & Handley, S. J. (2004). Belief bias and figural bias in syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 57*(4), 666-692. doi:10.1080/02724980343000440

Moutier, S., Plagne-Cayeux, S., Melot, A.-M., & Houde, O. (2006). Syllogistic Reasoning and Belief-Bias Inhibition in School Children: Evidence from a Negative Priming Paradigm. *Developmental Science, 9*(2), 166-172. doi:10.1111/j.1467-7687.2006.00476.x

Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research, 66*(3), 180-194.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: an individual differences approach. *Journal of Experimental Psychology: General, 125*(1), 4-27.

Simoneau, M., & Markovits, H. (2003). Reasoning with premises that are not empirically true: evidence for the role of inhibition and retrieval. *Developmental Psychology, 39*(6), 964.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3.

Stanovich, K. E. (2004). Metarepresentation and the great cognitive divide: A commentary on Henriques'"Psychology Defined". *Journal of clinical psychology, 60*(12), 1263-1266.

Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and Brain Sciences, 23*(05), 701-717.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). The complexity of developmental predictions from dual process models. *Developmental Review, 31*(2), 103-118.

Steegen, S., & De Neys, W. (2012). Belief inhibition in children’s reasoning: Memory-based evidence. *Journal of Experimental Child Psychology, 112*(2), 231-242.

Stupple, E. J., & Waterhouse, E. F. (2009). Negations in syllogistic reasoning: Evidence for a heuristic–analytic conflict. *The Quarterly Journal of Experimental Psychology, 62*(8), 1533-1541. doi:10.1080/17470210902785674

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. *two minds: dual processes and beyond. Oxford University Press, Oxford*.

Thompson, V. A. (2010). Towards a metacognitive dual process theory of conditional reasoning. *Cognition and conditionals: Probability and logic in human thinking*, 335-354.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127-154.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. e., ry. (2005). A dual-process specification of causal conditional reasoning. *Thinking \& Reasoning, 11*(3), 239-278.

Wason, P. C., & Evans, J. S. B. (1975). Dual processes in reasoning? *Cognition, 3*(2), 141-154.

Wetherick, N. (1989). Psychology and syllogistic reasoning. *Philosophical Psychology, 2*(1), 111-124.

Wetherick, N. E., & Gilhooly, K. J. (1995). Atompshere, matching, and logic in syllogistic reasoning. *Current Psychology, 14*(iss.3), 169-178.

Woodworth, R. S., & Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology, 18*(4), 451-460.

Chapter 8
General Discussion and Conclusions

Reasoning is a very complex cognitive process and it certainly requires the employment of cognitive resources. We found positive results for the involvement of working memory in all our four studies, for both preadolescent and adult participants. Working memory resources is essential for any effortful processing. As the central executive has limited capacity for both the storage and processing (Baddeley & Hitch, 1974). Processing demanding tasks would result in less resource for the storage, and vice versa. Thus, more difficult tasks, in terms of processing and/or storage, are more prone to error. Reasoning is one of the processing demanding tasks, especially the three-model problems. Besides working memory, the other cognitive capacities like the fluid intelligence, comprehension skills and inhibitory control are important as well. Our results are in line with mental model theory and dual-process theory's view that WM (in adults) and the development of this capacity (in preadolescents) is an important factor of reasoning performances, especially for difficult problems.

Effect of the grade and cognitive capacities on reasoning for adolescents

In the first study with preadolescents, we found reliable effect of verbal working memory in reasoning performances (propositional and syllogistic reasoning). However, in the second study with more cognitive measures, fluid intelligence was the most important predictor of reasoning performances for the first grader but the working memory capacities (visuospatial, verbal and central executive) for the second graders. Regression analysis showed that the fluid intelligence was the most important predictor of the reasoning performances (except syllogistic conclusion generation task) for the first grader but the overall WM capacities for the second graders. This supported the idea that the development of WM in preadolescence helps in the development of reasoning abilities. The importance of the development of different capacities in reasoning shows a rather complex pattern, in line with previous studies. The results pointed to theories which suggest the involvement of WM in reasoning, like the mental model theory and dual process theory assert.

Better reasoning performance with the course of preadolescence may due to both the increase in working memory, executive functions or cognitive capacities and the experiences and formal training gain (as well as the motivation to solve a problem correctly) with the grade level. Developments in the respective cognitive resources (WM and executive functions, particularly inhibitory control) might be necessary to facilitate children's ability to inhibit the irrelevant background knowledge and the output of heuristic process; and switch to formal reasoning processes. This view is supported by our result that 1. children with greater WM capacity have a higher percentage of correct responses (study 1 and 2). WM was a significant predictor of reasoning performances for second graders but it was fluid intelligence for first graders (study 2).

Besides working memory, the development of the metacognitive capacity which facilitate analytic reasoning has an extra progression during this period as well (we found significant difference even for the increment of just one year). Full control of metacognitive processes involved in reasoning is acquired late.

Effect of working memory and inhibitory control on reasoning for adults

Study 3 and 4 investigated two other factors of reasoning, namely task effect and inhibitory control, in addition to working memory. For study 3, we confirmed again that cognitive processes and resources required for generation and evaluation task of syllogistic reasoning are different. Also, automatic heuristics (namely the atmosphere/matching heuristics) affect evaluation more but WM capacity is more important for generation task. However, in general, more difficult problems (i.e. syllogisms of more mental models) load more on WM. We found a higher correlation of the multiple-model problems than the single-model problems with the visuospatial WM measure in study 3 (despite of the controversial results of the correlation analysis of reasoning performance with WM measures in study 4). The results suggest a dual-process explanation of syllogistic reasoning. The effect of number of models of the syllogism in the generation task and the use of atmosphere/matching heuristics, particularly in the evaluation task, were also confirmed.

The importance of inhibitory control for solving conflict problems (when the output from heuristics disagree with the one from logic) is also confirmed in study 4. However, significant correlation with WM measures was found for no-conflict problems but not for conflict problems. This result has to be interpreted with caution. It is not possible that WM is *not* involved in solving conflict problems but does in no-conflict ones. It is very possible that solving the very difficult conflict problems (3-model) in our task requires more WM capacities beyond the limitation of participants.

Regarding the suggestions of the cognitive theories on reasoning, the difficulty of syllogisms depends on whether solving the problem requires the involvement of Type 2 processing or just Type 1 processing (as suggested by dual-process theories), as well as the number of mental model of the syllogisms (as in mental model theory). Our results of study 3 were in line with one proposal of the Rule-based PSYCOP model (Rips, 1994) that reasoners can employ some backward reasoning processes (particularly in the evaluation task). Different reasoning strategies and heuristics were employed according to the task demand. We found that conclusion generation task is more prone to premise-driven processing; while evaluation task is more prone to conclusion-driven processing. This is another aspect which credit extra individual differences in reasoning, apart from cognitive capacities.

Many cognitive theories have been developed to gain a deeper understanding of how humans reason. A few meta-analysis studies have evaluated the predictive power of different cognitive theories of empirical data. There is indisputable individual difference in reasoning. In a forthcoming study that 139 participants were asked to solve all the 64 syllogisms, we found a very fuzzy result regarding the percentage of congruency of their response to the prediction of different theories. Cognitive theories are descriptive theories of how humans may reason and some of them propose how humans may deviate from normative theories such as classical logic. A problem of current theories of reasoning is that they mainly focus on the “average” reasoner (the aggregated result of many reasoners). We could not found any (or very few) individual reasoner who fit exactly the prediction of the very “accurate” models such as mReasoner (see also Table 2 in Khemlani & Johnson-Laird, 2016). It is thus very important for a good cognitive theory of reasoning to take into account factors of individual differences such as working memory and use of strategy. We observe the importance of working memory and inhibitory control in reasoning for both preadolescents and adults (in all the four studies). Besides, individual difference in terms of motivation and thinking style (Newstead, Handley, Harley, Wright, & Farrelly, 2004), experimental design and task demand can affect also affect the difference in reasoning performance.

Reference

Baddeley, A., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation, 8,* 47-89.

Khemlani, S., & Johnson-Laird, P. (2016). *How people differ in syllogistic reasoning.* Paper presented at the Proceedings of the 36th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology Section A, 57*(1), 33-60.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*: MIT Press.

References

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569.

Anderson, N. (1990). Spatial pattern of recent sediment and diatom accumulation in a small, monomictic, eutrophic lake. *Journal of Paleolimnology, 3*(2), 143-160.

Aslan, A., & Bäuml, K.-H. T. (2011). Individual differences in working memory capacity predict retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 264.

Bacon, A. M., Handley, S., & Newstead, S. E. (2003). Individual differences in strategies for syllogistic reasoning. *Thinking & Reasoning, 9*(2), 133-168. doi:10.1080/13546780343000196

Bacon, A. M., Handley, S. J., Dennis, I., & Newstead, S. E. (2008). Reasoning strategies: The role of working memory and verbal-spatial ability. *European Journal of Cognitive Psychology, 20*(6), 1065-1086.

Baddeley, A. (1997). *Human memory: Theory and practice*: Psychology Press.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation, 8,* 47-89.

Bara, B. G., Bucciarelli, M., & Johnson-Laird, P. N. (1995). Development of syllogistic reasoning. *American Journal of Psychology, 108*(2), 157.

Barrouillet, P. (2011). Dual-process theories of reasoning: The test of development.

Developmental Review, 31(2), 151-179.

Barrouillet, P., Gauffroy, C., & Lecas, J. F. (2008). Mental models and the suppositional account of conditionals. *Psychological Review, 115*(3), 760.

Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: a time-based resource-sharing model account. *Developmental Psychology, 45*(2), 477.

Barrouillet, P., Grosset, N., & Lecas, J. F. (2000). Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition, 75*(3), 237-266.

Barrouillet, P., & Lecas, J. F. (1998). How can mental models theory account for content effects in conditional reasoning? A developmental perspective. *Cognition, 67*(3), 209-253.

Barrouillet, P., & Lecas, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning, 5*(4), 289-302.

Barrouillet, P., Markovits, H., & Quinn, S. (2002). Developmental and content effects in reasoning with causal conditionals. *Journal of Experimental Child Psychology, 81*(3), 235-248.

Begg, I., & Denny, J. P. (1969). Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning errors. *Journal of Experimental Psychology: , Memory, and Cognition, 81*(2), 351-354.

Bell, V. A., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science, 22*(1), 25-51. doi:10.1016/S0364-0213(99)80034-2

Braine, M. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review, 85*(1), 1.

Braine, M., & O'Brien, D. P. (1998). *Mental logic*: Psychology Press.

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences, 16*(2), 106-113.

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. *Variation in working memory*, 76-106.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science, 23*(3), 247-303. doi:10.1016/S0364-0213(99)00008-7

Buehner, M., Krumm, S., & Pick, M. (2005). Reasoning= working memory≠ attention. *Intelligence, 33*(3), 251-272.

Byrne, R. M. J., & Johnson-Laird, P. N. (2009). `If'and the problems of conditional reasoning. *Trends in Cognitive Sciences, 13*(7), 282-287.

Capon, A., Handley, S., & Dennis, I. (2003). Working memory and reasoning: An individual differences perspective. *Thinking & Reasoning, 9*(3), 203-244.

Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences, 32*(02), 121-138.

Case, R. (1985). Intellectual development: Orlando, FL: Academic Press.

Castañeda, L. E. G., & Knauff, M. (2013). *Individual Differences, Imagery and the Visual Impedance Effect.* Paper presented at the CogSci.

Chater, N., & Oaksford, M. (1999). The Probability Heuristics Model of Syllogistic Reasoning. *Cognitive Psychology, 38*(2), 191-258. doi:10.1006/cogp.1998.0696

Chiesi, F., Primi, C., & Morsanyi, K. (2011). Developmental changes in probabilistic reasoning: The role of cognitive capacity, instructions, thinking styles, and relevant knowledge. *Thinking & Reasoning, 17*(3), 315-350.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407.

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*(2), 163-183.

Copeland, D., & Radvansky, G. (2004). Working memory and syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 57*(8), 1437-1457. doi:10.1080/02724980343000846

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422-433.

De Groot, A. M. (1983). The range of automatic spreading activation in word priming. *Journal of Verbal Learning and Verbal Behavior, 22*(4), 417-436.

De Neys, W. (2006). Automatic–heuristic and executive–analytic processing during reasoning: Chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology, 59*(6), 1070-1100.

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition, 113*(1), 45-61.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005a). Working memory and counterexample retrieval for causal conditionals. *Thinking & Reasoning, 11*(2), 123-150.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005b). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning, 11*(4), 349-381.

De Neys, W., & Van Gelder, E. (2009). Logic and Belief across the Lifespan: The Rise and Fall of Belief Inhibition during Syllogistic Reasoning. *Developmental Science*, *12*(1), 123-130. doi:10.1111/j.1467-7687.2008.00746.x

Dickstein, L. S. (1978). The effect of figure on syllogistic reasoning. *Memory and Cognition, 6*(1), 76-83.

Dietz, E.-A. (2015). A computational logic approach to syllogisms in human reasoning. *Bridging the Gap between Human and Automated Reasoning*, 17-31.

Elosúa, M. R., Carriedo, N., & García-Madruga, J. A. (2009). Dos nuevas pruebas de Memoria Operativa de Anáforas. *Infancia y aprendizaje*, *32*(1), 97-118.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309.

Espino, O., Santamaría, C., & García-Madruga, J. A. (2000). Figure and difficulty in syllogistic reasoning. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*.

Evans, J. S. B. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75*(4), 451-468.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10).

Evans, J. S. B. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*(3), 378-395.

Evans, J. S. B. (2007). On the resolution of conflict in dual process theories of reasoning.

Thinking & Reasoning, 13(4), 321-339.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol., 59*, 255-278.

Evans, J. S. B. (2009). How many dual-process theories do we need? One, two, or many?

Evans, J. S. B. (2010). Intuition and reasoning: A dual-process perspective. *Psychological Inquiry, 21*(4), 313-326.

Evans, J. S. B. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review, 31*(2), 86-102.

Evans, J. S. B., Barston, J., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition, 11*(3), 295-306. doi:10.3758/BF03196976

Evans, J. S. B., Handley, S. J., & Harper, C. N. J. (2001). Necessity, possibility and belief: A study of syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 54*(3), 935-958. doi:10.1080/713755983

Evans, J. S. B., Handley, S. J., Harper, C. N. J., & Johnson-Laird, P. N. (1999). Reasoning about necessity and possibility: A test of the mental model theoriy of deduction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1495-1513.

Evans, J. S. B., & Over, D. E. (1996). Rationality in the selection task: Epistemic utility versus uncertainty reduction.

Evans, J. S. B., & Over, D. E. (2004). *If*: Oxford: Oxford University Press.

Evans, J. S. B., & Stanovich, K. E. (2013a). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science, 8*(3), 223-241.

Evans, J. S. B., & Stanovich, K. E. (2013b). Theory and metatheory in the study of dual processing: Reply to comments. *Perspectives on psychological science, 8*(3), 263-271.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. *The nature of intelligence*, 231-235.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, *34*(10), 906.

Fletcher, L., & Carruthers, P. (2012). Metacognition and reasoning. *Phil. Trans. R. Soc. B*, *367*(1594), 1366-1378.

Ford, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition*, *54*(1), 1-71. doi:10.1016/0010-0277(94)00625-U

García-Madruga, J. A. (1982). Un estudio sobre el efecto de la figura en el razonamiento silogístico. *Estudios de Psicología*, *3*(11), 23-32.

García-Madruga, J. A. (1983). Un modelo general sobre el razonamiento silogístico: Doble procesamiento y fase de comprobación con verificación. *Revista de Psicología General y Aplicada*.

García-Madruga, J. A. (1989). Comprensión e inferencia en el razonamiento silogístico. *Cognitiva, 2*(3), 323-350.

García-Madruga, J. A., Elosúa, M. R., Gil, L., Gómez-Veiga, I., Vila, J. Ó., Orjales, I., . . . Duque, G. (2013). Reading comprehension and working memory's executive processes: An intervention study in primary school students. *Reading Research Quarterly*, *48*(2), 155-174.

García-Madruga, J. A., Gutiérrez, F., Carriedo, N., Luzón, J. M., & Vila, J. O. (2007). Mental models in propositional reasoning and working memory's central executive. *Thinking & Reasoning*, *13*(4), 370-393.

García-Madruga, J. A., Santamaría, C., Moreno-Ríos, S., Vila, Gómez-Chacón, Orenes, &

Navarrete. (2009). *Prueba de razonamiento para secundaria (Reasoning test for secondary school)*.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3*(11), 410-419.

Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology, 70*(2), 177-194.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*(2), 177.

Gauffroy, C., & Barrouillet, P. (2009). Heuristic and analytic processes in mental models for conditionals: An integrative developmental theory. *Developmental Review, 29*(4), 249-282.

Gauffroy, C., & Barrouillet, P. (2011). The primacy of thinking about possibilities in the development of reasoning. *Developmental Psychology, 47*(4), 1000.

Geurts, B. (2003). Reasoning with quantifiers. *Cognition, 86*(3), 223-251. doi:10.1016/S0010-0277(02)00180-4

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review, 103*(4), 650.

Gilhooly, K., Logie, R., & Wynn, V. (2002a). Syllogistic reasoning tasks and working memory: Evidence from sequential presentation of premises. *Current Psychology, 21*(2), 111.

Gilhooly, K. J., Logie, R., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & Cognition, 21*(1), 115-124.

Gilhooly, K. J., Logie, R., & Wynn, V. (1999). Syllogistic reasoning tasks, working memory, and skill. *European Journal of Cognitive Psychology, 11*(4), 473-498.

Gilhooly, K. J., Logie, R., & Wynn, V. (2002b). Syllogistic reasoning tasks and working memory: Evidence from sequential presentation of premises. *Current Psychology, 21*(2), 111-120. doi:10.1007/s12144-002-1007-x

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*(06), 803-831.

Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. S. B. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning, 10*(2), 175-195.

Hardman, D. K., & Payne, S. J. (1995). Problem difficulty and response format in syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 48*(4), 945-975.

Hasher, L., Lustig, C., Zacks, R., Conway, A., Jarrold, C., & Kane, M. (2007). Inhibitory mechanisms and the control of attention. *Variation in working memory, 19*, 227-249.

Houdé, O. (1997). The problem of deductive competence and the inhibitory control of cognition.

Inhelder, B., & Piaget, J. (1955). *De la logique de l'enfant à la logique de l'adolescent*. Paris: Presses Univ. de France. (2e éd. 1970.).

Inhelder, B., & Piaget, J. (1964). *Classification and seriation*: Routledge & Kegan Paul.

Isaacs, E. B., & Vargha-Khadem, F. (1989). Differential course of development of spatial and verbal memory span: A normative study. *British Journal of Developmental Psychology, 7*(4), 377-380.

Janveau-Brennan, G., & Markovits, H. (1999). The development of reasoning with causal conditionals. *Developmental Psychology, 35*(4), 904.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*: Harvard University Press.

Johnson-Laird, P. N. (1989). Mental models. In M. I. Posner (Ed.), *The Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology, 50*(1), 109-135.

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences, 5*(10), 434-442.

Johnson-Laird, P. N. (2006). *How we reason*. Oxford: University Press.

Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition, 16*(1), 1-61.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*: Lawrence Erlbaum Associates, Inc.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review, 109*(4), 646.

Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111*(3), 640.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., & Legrenzi, M. S. (2000). Illusions in reasoning about consistency. *Science, 288*(5465), 531-532.

Johnson-Laird, P. N., Oakhill, J., & Bull, D. (1986). Children's syllogistic reasoning. *The Quarterly Journal of Experimental Psychology, 38*(1), 35-58.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment, 49*.

Khemlani, S., & Johnson-Laird, P. (2016). *How people differ in syllogistic reasoning.* Paper presented at the Proceedings of the 36th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.

Khemlani, S., & Johnson-Laird, P. N. (2012). Theories of the Syllogism: A Meta-Analysis. *Psychological Bulletin, 138*(3), 427-457. doi:10.1037/a0026841

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *J Exp Psych*, 1-17.

Klaczynski, P. A. (2004). A dual-process model of adolescent development: Implications for decision making, reasoning, and identity. *Advances in child development and behavior, 32*, 73-123.

Klaczynski, P. A. (2005). Metacognition and cognitive variability: A dual-process model of decision making and its development. *The development of judgment and decision making in children and adolescents*, 39-76.

Klaczynski, P. A., Schuneman, M. J., & Daniel, D. B. (2004). Theories of conditional reasoning: a developmental examination of competing hypotheses. *Developmental Psychology, 40*(4), 559.

Klauer, K. C., Musch, J., & Naumer, B. (2000). On Belief Bias in Syllogistic Reasoning. *Psychological Review, 107*(4), 852-884. doi:10.1037/0033-295X.107.4.852

Knauff, M., & Johnson-Laird, P. N. (2002). Visual imagery can impede reasoning. *Memory & Cognition, 30*(3), 363-371.

Knauff, M., & May, E. (2006). Mental imagery, reasoning, and blindness. *The Quarterly Journal of Experimental Psychology, 59*(1), 161-177.

Kuhn, D. (2000). Metacognitive development. *Current directions in psychological science, 9*(5), 178-181.

Kuhn, D., & Franklin, S. (2006). *The second decade: What develops (and how)*: Wiley Online Library.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*(4), 389-433.

Lustig, C., Hasher, L., & Zacks, R. T. (2007). Inhibitory deficit theory: Recent developments in a "new view". *Inhibition in cognition, 17*, 145-162.

Markovits, H. (2004). The development of deductive reasoning. *The nature of reasoning*, 313-338.

Markovits, H. (2014). On the road toward formal reasoning: Reasoning with factual causal and contrary-to-fact causal premises during early adolescence. *Journal of Experimental Child Psychology, 128*, 37-51.

Markovits, H., & Barrouillet, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review, 22*(1), 5-36.

Markovits, H., Doyon, C., & Simoneau, M. (2002). Individual differences in working memory and conditional reasoning with concrete and abstract content. *Thinking & Reasoning, 8*(2), 97-107.

Markovits, H., & Lortie-Forgues, H. (2011). Conditional reasoning with false premises facilitates the transition between familiar and abstract reasoning. *Child development, 82*(2), 646-660.

Markovits, H., & Nantel, G. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition*, *17*(1), 11-17.

Markovits, H., Schleifer, M., & Fortier, L. (1989). Development of elementary deductive reasoning in young children. *Developmental Psychology, 25*(5), 787.

Markovits, H., & Thompson, V. (2008). Different developmental patterns of simple deductive and probabilistic inferential reasoning. *Memory & Cognition, 36*(6), 1066-1078.

Markovits, H., & Vachon, R. (1989). Reasoning with contrary-to-fact propositions. *Journal of Experimental Child Psychology, 47*(3), 398-412.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227.

Miller, S. A., Custer, W. L., & Nassau, G. (2000). Children's understanding of the necessity of logically necessary truths. *Cognitive Development, 15*(3), 383-403.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*: Cambridge University Press.

Morley, N. J., Evans, J. S. B., & Handley, S. J. (2004). Belief bias and figural bias in syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 57*(4), 666-692. doi:10.1080/02724980343000440

Moshman, D. (1990). Rationality as a goal of education. *Educational Psychology Review, 2*(4), 335-364.

Moshman, D. (2004). From inference to reasoning: The construction of rationality. *Thinking & Reasoning, 10*(2), 221-239.

Moshman, D., & Franks, B. A. (1986). Development of the concept of inferential validity. *Child development*, 153-165.

Moutier, S., Plagne-Cayeux, S., Melot, A.-M., & Houde, O. (2006). Syllogistic Reasoning and Belief-Bias Inhibition in School Children: Evidence from a Negative Priming Paradigm. *Developmental Science, 9*(2), 166-172. doi:10.1111/j.1467-7687.2006.00476.x

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology*

Section A, 57(1), 33-60.

Neys, W. D. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological science, 17*(5), 428-433.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*(4), 608.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*: Oxford University Press.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(4), 883.

Oaksford, M., Roberts, L., & Chater, N. (2002). Relative informativeness of quantifiers used in syllogistic reasoning. *Memory & Cognition, 30*(1), 138-149. doi:10.3758/BF03195273

Orjales, I., García-Madruga, J. A., & Elosúa, M. (2010). Prueba de amplitud lectora para primaria [Reading span test for primary school]. *Unpublished manuscript*.

Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research, 66*(3), 180-194.

Piaget, J. (1953). The origins of intelligence in children. *Journal of Consulting Psychology, 17*(6), 467.

Piaget, J. (1970). Science of education and the psychology of the child. Trans. D. Coltman.

Piaget, J. (1987). *Possibility and necessity* (Vol. 1): U of Minnesota Press.

Pillow, B. H. (2002). Children's and adults' evaluation of the certainty of deductive inferences, inductive inferences, and guesses. *Child development, 73*(3), 779-792.

Quayle, J. D., & Ball, L. J. (2000). Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A, 53*(4), 1202-1223. doi:10.1080/713755945

Revlis, R. (1975). Two models of syllogistic inference: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior, 14*(2), 180-195.

Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review, 90*(1), 38.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*: MIT Press.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological Bulletin, 126*(1), 3.

Roberts, M. J., & Sykes, E. D. (2005). Categorical reasoning from multiple diagrams. *The Quarterly Journal of Experimental Psychology Section A, 58*(2), 333-376.

Ruff, C. C., Knauff, M., Fangmeier, T., & Spreer, J. (2003). Reasoning and working memory: common and distinct neuronal processes. *Neuropsychologia, 41*(9), 1241-1253.

Santamaría, C., Tse, P. P., Moreno-Ríos, S., & García-Madruga, J. A. (2013). Deductive reasoning and metalogical knowledge in preadolescence: a mental model appraisal. *Journal of Cognitive Psychology, 25*(2), 192-200.

Schroyens, W. (2010). A meta-analytic review of thinking about what is true, possible, and irrelevant in reasoning from or reasoning about conditional propositions. *European Journal of Cognitive Psychology, 22*(6), 897-921.

Schroyens, W., Schaeken, W., Fias, W., & d'Ydewalle, G. e., ry. (2000). Heuristic and analytic processes in propositional reasoning with negatives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1713-1734.

Sells, S. B. (1936). The atmosphere effect: an experimental study of reasoning. *Archives of Psychology (Columbia University)*.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: an individual differences approach. *Journal of Experimental Psychology: General*, *125*(1), 4-27.

Sima, J. F., Schultheis, H., & Barkowsky, T. (2013). Differences between spatial and visual mental representations. *Frontiers in psychology, 4*(240). doi:http://doi.org/10.3389/fpsyg.2013.00240

Simoneau, M., & Markovits, H. (2003). Reasoning with premises that are not empirically true: evidence for the role of inhibition and retrieval. *Developmental Psychology, 39*(6), 964.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and social psychology review, 4*(2), 108-131.

Stanovich, K. E. (2004). Metarepresentation and the great cognitive divide: A commentary on Henriques'"Psychology Defined". *Journal of clinical psychology, 60*(12), 1263-1266.

Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought*: Yale University Press.

Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and Brain Sciences, 23*(05), 701-717.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). The complexity of developmental predictions from dual process models. *Developmental Review, 31*(2), 103-118.

Steegen, S., & De Neys, W. (2012). Belief inhibition in children's reasoning: Memory-based evidence. *Journal of Experimental Child Psychology, 112*(2), 231-242.

Stenning, K., & Lambalgen, M. (2008). *Human reasoning and cognitive science*. Cambridge, MA: MIT Press.

Störring, G. (1908). *Experimentelle untersuchungen über einfache Schlussprozesse*: W. Engelmann.

Stupple, E. J., Ball, L., & Ellis, D. (2013). Matching bias in syllogistic reasoning: Evidence for a dual-process account from response times and confidence ratings. *Thinking & Reasoning, 19*(1), 54-77. doi:10.1080/13546783.2012.735622

Stupple, E. J., & Ball, L. J. (2005). *An inspection-time analysis of figural effects and processing direction in syllogistic reasoning.* Paper presented at the Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society.

Stupple, E. J., & Ball, L. J. (2007). Figural effects in a syllogistic evaluation paradigm: An inspection-time analysis. *Experimental Psychology, 54*(2), 120-127.

Stupple, E. J., & Waterhouse, E. F. (2009). Negations in syllogistic reasoning: Evidence for a heuristic–analytic conflict. *The Quarterly Journal of Experimental Psychology, 62*(8), 1533-1541. doi:10.1080/17470210902785674

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence, 30*(3), 261-288.

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. *two minds: dual processes and beyond. Oxford University Press, Oxford*.

Thompson, V. A. (2010). Towards a metacognitive dual process theory of conditional reasoning.

Cognition and conditionals: Probability and logic in human thinking, 335-354.

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, *128*(2), 237-251.

Toms, M., Morris, N., & Ward, D. (1993). Working memory and conditional reasoning. *The Quarterly Journal of Experimental Psychology, 46*(4), 679-699.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127-154.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1130. doi:10.1126/science.185.4157.1124

Velasco, J., & García-Madruga, J. A. (1997). La búsqueda de contraejemplos como estrategia metalógica. Un estudio sobre la influencia de su entrenamiento en el razonamiento silogístoco. *Infancia y aprendizaje*, *20*(80), 17-35.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. e., ry. (2005). A dual-process specification of causal conditional reasoning. *Thinking \& Reasoning, 11*(3), 239-278.

Wason, P. C., & Evans, J. S. B. (1975). Dual processes in reasoning? *Cognition, 3*(2), 141-154.

Wetherick, N. (1989). Psychology and syllogistic reasoning. *Philosophical Psychology, 2*(1), 111-124.

Wetherick, N. E., & Gilhooly, K. J. (1995). Atompshere, matching, and logic in syllogistic reasoning. *Current Psychology, 14*(iss.3), 169-178.

Whitney, P., Arnett, P. A., Driver, A., & Budd, D. (2001). Measuring central executive functioning: What's in a reading span? *Brain and Cognition, 45*(1), 1-14.

Wilson, J., Scott, J. H., & Power, K. G. (1987). Developmental differences in the span of visual memory for pattern. *British Journal of Developmental Psychology, 5*(3), 249-255.

Woodworth, R. S., & Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology, 18*(4), 451-460.

Zhai, F. (2015). *Toward Probabilistic Natural Logic for Syllogistic Reasoning* (Unpublished master thesis). Universiteit van Amsterdam, Amsterdam, Netherlands.

Zhai, F., Szymanik, J., & Titov, I. (2015). *Toward a probabilistic mental logic for the syllogistic fragment of natural language.* Paper presented at the Proceedings of the 20th Amsterdam Colloquium.

Appendix A

The 18 reasoning problems used in the questionnaire

(* = correct response)

Task 1. Propositional inference task

1. (DA) For the first dish, the teacher says:

"If I use mayonnaise, then I use lettuce"

And then asserts:

"I do not use mayonnaise"

What can you conclude about the other ingredient?

-They use lettuce

-They do not use lettuce

-Cannot form a conclusion.*

2. (MT) For the second dish, the teacher says:

"If I use garlic then I use oil"

And then asserts:

"I do not use oil"

What can you conclude about the other ingredient?

-They use garlic

-They do not use garlic*

-Cannot form a conclusion.

3. (Affirmative disjunction) For the first dessert, the teacher says:

"I use honey or cinnamon, or both"

And then asserts:

"I use honey"

What can you conclude about the other ingredient?

-They use cinnamon

-They do not use cinnamon

-Cannot form a conclusion.*

4. (Negative disjunction) For the second dessert, the teacher says:

"I use chocolate or wine, or both"

And then asserts:

"I do not use wine"

What can you conclude about the other ingredient?

-They use chocolate*

-They do not use chocolate

-Cannot form a conclusion.

Task 2. Evaluation of propositional attitudes

1. Sonia said she would go swimming or to soccer, but not both.

Finally she did not go swimming, but she went to soccer.

Did Sonia tell the truth? Yes* / No

2. Andrés said he would go to typing-class or English-class, but not both.

Finally he didn’t go to typing-class and he didn’t go to English-class.

Did Andrés tell the truth? Yes / No*

3. Marta said that if she went to basketball then she would not go to tennis.

Finally she went to basketball and tennis.

Did Marta tell the truth? Yes / No*

4. Paco said that if he went to chess then he would go to squash.

Finally he did not go to chess and he did not go to squash.

Did Paco tell the truth? Yes* / No

5. Laura said that if she went to karate then she would go to the gym

Finally she didn’t go to karate but she went to the gym.

Did Laura tell the truth? Yes* / No

Task 3. Syllogistic construction task

1. All skiers are vegetarians.

Some plumbers are skiers

Conclusion?

2.Some mechanics are teachers

All motorists are mechanics

Conclusion?

3. No football player is a programmer.

All football players are athletes

Conclusion?

Task 4. Modal syllogistic task

1. All engineers are mathematicians

Sergio is an engineer

Conclusion: Sergio is not a mathematician

Necessarily True / Possible / Impossible*

2. All artists are gardeners

Raquel is an artist

Conclusion: Raquel is a gardener

Necessarily True*/ Possible / Impossible

3. All doctors are physicists

Luis is a physicist

Conclusion: Luis is a doctor

Necessarily True / Possible* / Impossible

Task 5. Probabilistic reasoning task

1. There were two hospitals in a city. Around 135 babies are born in the larger hospital every day and 30 babies in the smaller hospital every day. As you know, around 50% of the newborns are males. However, the exact percentage varies each day. It is larger than 50% sometimes and smaller than 50% sometimes. In a year, each hospital has registered the days in which 85% of the newborns were males. Which of the hospitals do you think has registered more of such days?

a) The larger

*b) The smaller

c) Both hospitals have registered them with the same frequency

2. We have passed a survey in all the families with six children in a village in Jaén. The parents have to note the exact order of the birth of the boys (V) and girls (M). Which of the following sequence, MMVMVV or VVVVMV, do you think is more possible that the González Pérez family would note?

a) the first

b) the second

*c) Both has the same probability

3. A test which can diagnose the cancer was performed by all the residents in a big city in which

there are few cancer cases. Evidently, like all other tests, it gives false positive results occasionally. That means the test indicate that the patient has cancer thought he does not have it in reality.

Which of the following is more possible for you?

a) A person has cancer if the test has shown a positive result

*b) The test shows positive result if the person has cancer.

c) Both events have the same probability.

Appendix B

Atmosphere and validity of the conclusions.

	conclusion is consistent with atmosphere	conclusion is NOT consistent with atmosphere
conclusion is valid	P1: All B are A P2: Some B are C C: **Some A are C**	P1: All B are A P2: All B are C C: **Some A are C**
conclusion is invalid	P1: All B are A P2: No B are C C: *No A are C (**Some A are not C**)	P1: No A are B P2: All B are C C: *Some A are C (**Some C are not A**)

Note. Valid conclusions are in bold.

Appendix C

Number and type of problems used in the evaluation task.

	single-Atm	single-NoAtm	multiple-Atm	multiple-NoAtm
valid	9	0	12	6
invalid	1	8	12	6

Appendix D

List of the 27 syllogistic problems used in generation and evaluation tasks, including mood, figure, number of mental models and percentages of correct conclusions.

Problems	Generation task	Evaluation task		
		Valid	Invalid	Overall
1. IA-I (1) 1m	85%	79% (I-AC)	*79% (E-CA)*	79%
2. AI-I (2) 1m	80%	80% (I-CA)	*68% (E-CA)*	74%
3. EI-O (1) 3m	15%	63% (O-CA)	42% (O-AC)	53%
4. IE-O (2) 3m	28%	86% (O-AC)	35% (O-CA)	62%
5. EI-O (2) 3m	45%	45% (O-CA)	41% (O-AC)	43%
6. IE-O (1) 3m	73%	79% (O-AC)	47% (O-CA)	63%
7. IA-I (4) 1m	60%	86% (I-AC)	*85%* (*A-CA)*	86%
8. AI-I (4) 1m	66%	84% (I-CA)	*74% (E-CA)*	79%
9. OA-O (3) 2m	23%	65% (O-AC)	*27% (I-CA*)	45%
10. AO-O (3) 2m	32%	74% (O-CA)	*32% (I-CA)*	53%
11. AA-A (2) 1m		66% (A-CA)	28% (A-AC)	48%
12. AA-I[20] (1) 3m		*41% (I-CA)*	22% (A-CA)	31%
13. AA-I (4) 3m		*39% (I-AC)*	21% (A-AC)	30%
14. AE-E (3) 1m		64% (E-AC)	*54% (I-CA)*	59%
15. EA-E (3) 1m		71% (E-CA)	*64% (I-CA)*	68%
16. AE-O (2) 3m		*56% (O-AC)*	32% (E-CA)	44%
17. EA-O (1) 3m		*39% (O-CA)*	*59% (I-AC)*	49%

[20] For syllogism 12, A-AC is also a valid conclusion but we have presented the no-Atm I-CA valid conclusion.

18. EA-O (4) 3m		*49% (O-CA)* *44% (O-AC)* 46%
19. AE-O (4) 3m		*51% (O-AC)* 23% (E-AC) 38%
20. AE-E (1) 1m		46% (E-CA) *73% (I-CA)* 60%
21. EA-E (2) 1m		66% (E-CA) *69% (A-CA)* 68%
22. IE-O (4) 3m		82% (O-AC) 37% (O-CA) 59%
23. EI-O (4) 3m		76% (O-CA) 23% (O-AC) 50%
24. IE-O (3) 3m		56% (O-AC) 34% (O-CA) 45%
25. EI-O (3) 3m		59% (O-CA) *54% (E-AC)* 56%
26. OA-O (4) 2m		77% (O-CA) *56% (E-CA)* 66%
27. AO-O (4) 2m		83% (O-AC) 33% (O-CA) 59%

Note. In IA-I (1) 1m, the letters refer to the mood of the premises and conclusion, and the number in parenthesis refers to their figure; 1m = single-model; 2m = 2-model; 3m = 3-model. A-AC = All As are Cs; A-CA = All Cs are As; I-AC = Some As are Cs; I-CA = Some Cs are As; O-AC = Some As are not Cs; O-CA = Some Cs are not As; E-AC = No A is C; E-CA = No C is A.

Appendix E

Figure codes of the syllogisms

figure 1	figure 2	figure 3	figure 4
A-B	B-A	A-B	B-A
B-C	C-B	C-B	B-C

Appendix F

Valid conclusions of all possible syllogisms

mood/figure	1-AB/BC	2-BA/CB	3-AB/CB	4-BA-BC
AA	***Aac**, Iac, Ica (1)	***Aac**, Iac, Ica (1)		Iac, Ica (3)
AI		***Iac,*Ica** (1)		***Iac, *Ica** (1)
AO			***Oca** (2)	***Oac** (2)
AE	***Eac, *Eca**, Oac, Oca (1)	Oac (3)	***Eac, *Eca**, Oac, Oca (1)	Oac (3)
IA	***Iac, *Ica** (1)			***Iac, *Ica** (1)
II				
IO				
IE	Oac (3)	Oac (3)	Oac (3)	Oac (3)
OA			***Oac** (2)	***Oca** (2)
OI				
OO				
OE				
EA	Oca (3)	***Eac, *Eca**, Oac,Oca (1)	***Eac, *Eca**, Oac,Oca (1)	Oca (3)
EI	Oca (3)	Oca (3)	Oca (3)	Oca (3)
EO				
EE				

*agree with matching heuristic.

Note. Number of mental models in parentheses. The valid conclusions of IE and EI syllogisms agree with atmosphere but not the matching heuristic.

Appendix G

List of the syllogisms used and their respective related and unrelated words used in the lexical decision task (translated from Spanish)

Problem type	Conclusion type	Syllogisms	Syllogisms in the experiment
8 conflict problems (multiple-model)	4 match-invalid	2 AE-E(2)*	All flowers need water No computer is a flower No computer needs water RW: Water, thirst, sea, computer, screen, keyboard UW: wolf, date, tie, cooked, firm, annoying
			All snakes are reptiles No rabbit is a snake No rabbit is a reptile RW: reptile, scales, crocodile, rabbit, hunting, gun UW: nose, beer, plum, euro, thunder, lamp
		2 EA-E(1)	No elephant is a car All cars need petrol No elephant needs petrol RW: elephant, giant, tube, petrol, truck, drive UW: flexible, rattle, ax, head, enough, history
			No stone is a cat All cats have eyes No stone has eyes RW: stone, rock, sand, eyes, eyelashes, pupils UW: cup, fashion, bath, cinnamon, terrace, lake
	4 Mismatch-valid	2 AE-O(2)	All dolphins are animals No cold-blooded animal is a dolphin Some animals are not cold-blooded RW: animal, life, natural, cold-blooded, turtle, fish UW: living room**, pepper, kilogram, sadness, mountain, sofa

			All pilots are drivers Nobody with autism is a pilot Some drivers are not people with autism RW: driver, chauffeur, taxi, autism, alone, afraid UW: soil, chemistry, orange, diamond, rivers, daisy
		2 EA-O(1)	All creams are objects with lots of odour No object with lots of odour is metallic Some metallic objects are not creams RW: cream, cosmetics, face, metallic, gold, iron UW: pistol, wine, rain, wool, tower, war
			All sailing champions swim well Nobody who swims well is paralytic Some people who are paralytic are not sailing champions RW: sailing, travel, sailing, paralysis, stroke, stationary UW: knife, shirt, hen, cement, hail, grass
6 No-conflict problems (single-model)	4 Match-valid	2 AE-E(1)	All ants are insects No insect has a backbone No ant has a backbone RW: ant, insect, cockroach, vertical column, spine, bone UW: hole, joke, spoon, drawing, room, video
			Finland is in the north No country in the north is hot Finland is not hot RW: Finland, forest, snow, heat, sweat, sun UW: hall, abdomen, stockings, smart, alphabet, tender
		2 EA-E(2)	No fruit is a dairy product All apples are fruits No apple is a dairy product RW: milk, cheese, beef, apple, pear, fruit UW: pants, telephone, shoes, island, cloud, idol

			No animal is immortal All birds are animals No bird is immortal RW: immortal, soul, angel, bird, wings, fly (verb) UW: hole, joke, spoon, drawing, room, video
	2 Mismatch-invalid	AE-O(1)	All men have the Y-gene Nobody with the Y-gene has ova Some men do not have ova RW: man, male, father, eggs, women, pregnancy UW: fly (insect), skyscrapers, table, always, shank, brick
		EA-O(2)	No object with wheels is a boat All vehicles have wheels Some vehicles are not boats RW: boats, cruise, ocean, vehicle, car, van UW: cotton, guitar, bend, arm, skirt, parsley

Note. RW: Related words; UW: Unrelated words. *Figure in parenthesis. ** Living room is "salon" in Spanish (a single word).