Deduction without awareness

Carlo Reverberi[1,2]*, Doris Pischedda[1], Michele Burigo[3], Paolo Cherubini[1]*

[1] Department of Psychology, University of Milano-Bicocca, 20126 Milan, Italy

[2] Bernstein Center for Computational Neuroscience Berlin, Charité-Universitätsmedizin, 10115 Berlin, Germany

[3] Language and Cognition Group, Cognitive Interaction Technology (CITEC), 33615 Bielefeld, Germany

*Corresponding authors:

Department of Psychology

University of Milano-Bicocca

Piazza dell'Ateneo Nuovo, 1

20126 Milan, Italy

E-mail: carlo.reverberi@unimib.it (C. Reverberi); paolo.cherubini@unimib.it (P. Cherubini)

Abstract

We investigated whether two basic forms of deductive inference, Modus Ponens and Disjunctive Syllogism, occur automatically and without awareness. In Experiment 1, we used a priming paradigm with a set of conditional and disjunctive problems. For each trial, two premises were shown. The second premise was presented at a rate designed to be undetectable. After each problem, participants had to evaluate whether a newly-presented target number was odd or even. The target number matched or did not match a conclusion endorsed by the two previous premises. We found that when the target matched the conclusion of a Modus Ponens inference, the evaluation of the target number was reliably faster than baseline even when participants reported that they were not aware of the second premise. This priming effect did not occur for any other valid or invalid inference that we tested, including the Disjunctive Syllogism. In Experiment 2, we used a forced-choice paradigm in which we found that some participants were able to access some information on the second premise when their attention was explicitly directed to it. In Experiment 3, we showed that the priming effect for Modus Ponens was present also in subjects who could not access any information about P2. In Experiment 4 we explored whether spatial relations (e.g., "a before b") or sentences with quantifiers (e.g., "all a with b") could generate a priming effect similar to the one observed for Modus Ponens. A priming effect could be found for Modus Ponens only, but not for the other relations tested. These findings show that the Modus Ponens inference, in contrast to other deductive inferences, can be carried out automatically and unconsciously. Furthermore, our findings suggest that critical deductive

inference schemata can be included in the range of high-level cognitive activities that are carried out unconsciously.

1. Introduction

The ability to recognise simple deductive arguments is a core property of rational thought (Rips, 1988). Two of the most intuitively straightforward deductive inferences are the *Modus Ponens* and the *Disjunctive Syllogism*:

1. Modus Ponens: If A then B; A; Therefore, B;

2. Disjunctive Syllogism: A or B or both; Not A; Therefore, B.

Acceptance of the Modus Ponens inference is nearly ubiquitous, with individuals correctly applying the inference in roughly 90% to 100% of the cases. Disjunctive Syllogism inferences are slightly more difficult and are, on average, correctly applied 80% of the times (Braine & O'Brien, 1998; Evans, Newstead, & Byrne, 1993). Because these two deductive processes are so fundamental to human rationality, some scholars contend that Modus Ponens and Disjunctive Syllogism inferences occur automatically and are triggered whenever a set of premises matches the corresponding logical schema, even without intention (Braine & O'Brien, 1998). Thus far, this issue has not been definitively settled by empirical studies. Nevertheless, some methodologically interesting strategies have been proposed, which capitalise on "priming effects" or the effects of pre-activated information on the processing of target stimuli (Meyer & Schvaneveldt, 1971; Neely, 1977). In the present case, if, for example, Modus Ponens were fully automatic, given the premise pair: "If there is a 3 then there is an 8" (major premise, P_1) and "there is a 3" (minor premise, P_2), individuals should activate the conclusion "there is an 8", even if they did not intentionally try to complete this premise pair. Similarly, if the Disjunctive Syllogism is automatic, the pair of premises "there is not a 3 or there is an 8" (P_1)

and "there is a 3" (P_2) should unavoidably activate "there is an 8". If those inferences are automatically triggered, there should be a priming effect when participants are shown, after the premise pair, a target number (T) that matches the conclusion ("8" in the previous examples), even if the participant has not been required to deduce a conclusion from the premises. In this way it would be possible to examine the automaticity of deductive reasoning across different types of inferences. However, all previous investigations of priming effects in processing of Modus Ponens or Disjunctive Syllogisms have used text comprehension tasks, in which the premises are embedded in text that participants are explicitly required to understand (Lea, 1995; Rader & Sloutsky, 2002). Under these circumstances, it is difficult to argue that participants are not voluntarily reasoning, even though they are not explicitly told to do so. Subliminally presented stimuli can be processed at a semantic level, as demonstrated by Naccache and Dehaene (2001b). Hence, it is plausible that subliminally presented stimuli may trigger inferences, if another premise is already encoded in WM and if that inference can be made without voluntary control. The finding that valid conclusions in Modus Ponens or Disjunctive Syllogism problems prime subsequent targets when one of the problem premises is not consciously available to the participant may indicate that the inference was triggered and carried out without voluntary control.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Fifty-four (44 females) graduate and undergraduate students from the University of Milano-Bicocca participated in the experiment in exchange for course credit.

2.1.2. Stimuli

The experiment involved two tasks: a *number evaluation task* (512 trials), designed to detect priming effects on target numbers, and a *sentence recognition task* (128 trials). The aim of the sentence recognition task was to check whether the subjects correctly encoded the first premise (P_1) in memory (Fig. 1), as explicitly instructed (see below), and to exclude from the analyses those who did not. All trials had two premises, P_1 and P_2. In the number evaluation trials, the second premise (P_2) was followed by a target number (T) and participants were required to judge whether the number was odd or even. In the sentence recognition trials, P_2 was followed by two sentences, one of them identical to the P_1. Participants were required to choose the sentence identical to P_1.
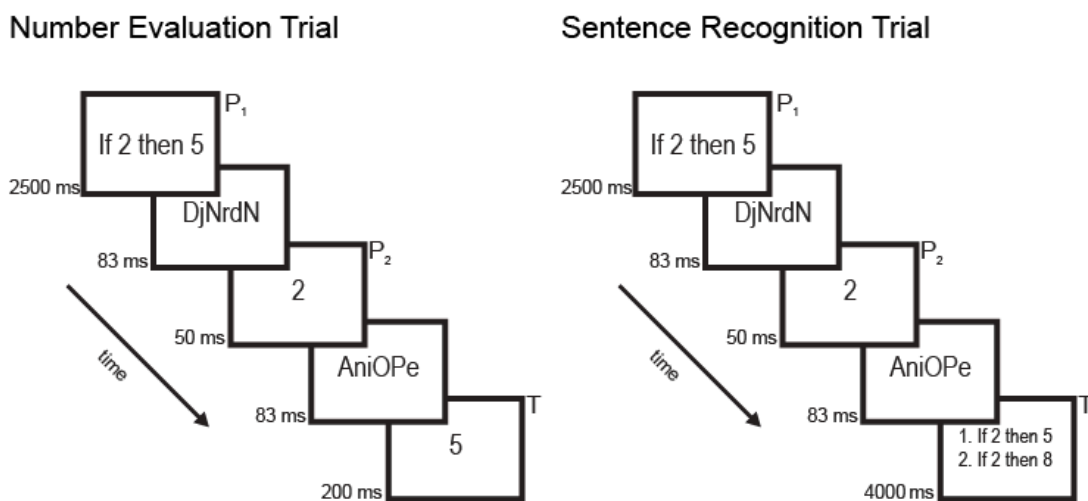

Figure 1. An example of two trial types used in Experiments 1-3. On the left, we show an example of a number evaluation trial. The participants had to read and remember the first premise (P_1).

Following P_1, P_2 was presented very briefly (50ms), preceded and followed by random string of ten characters. This masking procedure prevented most participants (see Experiments 2 and 3) to consciously access the number presented on P_2. Finally, a target number (T) was presented briefly (200ms). The participants had to judge within 2000ms from target presentation whether the number was odd or even. Sentence recognition trials (example on the right) were very similar to number evaluation trials. The only difference was at the target phase: instead of a number, two sentences were presented. The participant had to recognise the sentence that was identical to P_1. The target sentences remained on the screen for at most 4000ms.

More specifically, in number evaluation trials, P_1 was either a conditional (e.g., "if there is a 2 then there is a 4") or a disjunctive (e.g., "there is a 2 or there is a 4") statement. P_2 and T were numbers (e.g., "2"). As a short code for identifying each experimental trial in the following paragraphs, we state the P_1, P_2, and target number in a synthetic form. For instance, "if p then not q; r; p" stands for "if there is a p, then there is not a q (P_1); r (P_2); p (T)". The literals p, q, r, and s stand for Arabic numerals. Matching literals correspond to matching numbers. During the experiment, the specific numbers used as p, q, r, or s were randomly chosen from numbers between 1 and 8. Eight different types of P_1 were administered, corresponding to the cells of a 2 (logical connective: if *vs.* or) × 4 (negation: none, first argument, second argument, both arguments) design. P_2 was either identical to one of the numbers in P_1 (i.e., p, q) or different from both of them (r). The target (T) could be p, q, r, or s, being s a number different from all the preceding ones. Overall, 64 trial types were considered. The trial types relevant for the assessment of the automaticity of inferences were 32 (Table 1). The other trials were used as fillers. In the filler trials either T was identical to P_2 (e.g., "p or q; p; p") or the numbers involved in the whole trial were different (e.g., "p or q; r; s").

Table 1. Trial types used for assessing inference making.

	Inference Problem	**Baseline**
Conditional Problems		
1.	If p then q; p; q	If p then q; r; q
2.	If not p then q; p; q	If not p then q; r; q
3.	If p then not q; p; q	If p then not q; r; q
4.	If not p then not q; p; q	If not p then not q; r; q
5.	If p then q; q; p	If p then q; r; p
6.	If not p then q; q; p	If not p then q; r; p
7.	If p then not q; q; p	If p then not q; r; p
8.	If not p then not q; q; p	If not p then not q; r; p
Disjunctive Problems		
9.	p or q; p; q	p or q; r; q
10.	not p or q; p; q	not p or q; r; q
11.	p or not q; p; q	p or not q; r; q
12.	not p or not q; p; q	not p or not q; r; q
13.	p or q; q; p	p or q; r; p
14.	not p or q; q; p	not p or q; r; p
15.	p or not q; q; p	p or not q; r; p
16.	not p or not q; q; p	not p or not q; r; p

The literals p, q, r, and s stand for Arabic numerals between 1 and 8. Matching literals correspond to matching numbers.

Overall, the design allowed that the first premises (P_1) could or could not endorse a valid conclusion, depending on the presence of negations, the type of connective, and the type of P_2. Beyond Modus Ponens and Disjunctive Syllogism inferences, the problems comprised a wide array of other inferences, such as valid *Modus Tollens* ("if A then B; not B; therefore, not A", tested in trials such as "if not p then not q; q; p") and four invalid inferences that are often endorsed by naïve reasoners: "*Affirmation of the Consequent*" ("if A then B; B; therefore, A"), "*Denial of the Antecedent*" ("if A then B; not A; therefore, not B") and the corresponding invalid disjunctive inferences "A or B; A; therefore, not B" and "A

or B; B; therefore, not A". Each experimental condition was matched with a corresponding *baseline* condition in which all the elements visible to the subject (P_1 and T) were identical to those in the experimental condition but with a different P_2, not visible to the participants (see Table 1). Thus, for example, in the case of the Modus Ponens trials ("if p then q; p; q"), the relevant baseline was ("if p then q; r; q"). Given the critical difference in P_2 between experimental and baseline conditions, any inference possibly triggered in the experimental trials would not be triggered in the corresponding baseline trials. Importantly, however, any effect related only to P_1 and T (the visible part of the problems) would be equally present in both the experimental and baseline conditions. In this way, effects related only to the interaction between P_1 and T would be cancelled out.

2.1.3. Procedure

Participants were informed that they would be presented with sentences about the possible presence of digits on an imaginary blackboard. There were no explicit limits on the number of digits that the blackboard might contain. Following this, stating "There is a 2" did not imply the absence of other digits than 2. Each trial started with a fixation point (1000ms), followed by the presentation of P_1 for 2500ms. Participants had to read and remember P_1. After P_1, a mask of 10 random characters was displayed for 83ms (Fig. 1), followed by the presentation of P_2 (50ms), and then by another 10-character mask (83ms). After the second mask, there was a delay of 750ms in the long inter-stimulus interval (ISI) group and no delay for the short ISI group. This method was done to sample priming effects across different time frames (Draine & Greenwald,

1998). In number evaluation trials, the target number was presented for 200ms and the participant had to determine whether the number was odd or even. Up to 2000ms were allowed for parity judgment on the target. If no response was produced within the time limit, the trial ended and was scored as incorrect. In sentence recognition trials, a pair of statements was presented and the participants had to identify the one that exactly matched P_1. The statements disappeared as soon as the participants responded or after 4000ms if no response was given. The two response keys were the same in both tasks, one of them meaning "first sentence" or "odd number" and the other meaning "second sentence" or "even number". The response mapping was counterbalanced across subjects.

Before the experiment, participants were given written instructions describing the task. The presence of a number (P_2) between the masks was not mentioned. The importance of answering both quickly and correctly was stressed in the number evaluation trials. For sentence recognition trials, accuracy was emphasised more than speed. A 40-trial training session with accuracy feedback was administered before the experimental session. RTs on the number evaluation task were used to assess whether processing of targets was affected by the deductive conclusions entailed by the premises. Subjects with an average score lower than 90% on the sentence recognition trials were discarded from further analyses because the triggering of an inference – if any – strictly depended on an adequate representation of P_1.

2.1.4.   Apparatus

The experiment was programmed and administered using Presentation[(R)] (Neurobehavioral Systems, Inc., San Francisco, CA, US). The CRT screen had a 60Hz refresh rate and the resolution was set at 1024 × 768 pixels. The premises, masks, and target stimuli were written in white, 22-point Arial characters and were presented against a black background. Two buttons on a serial mouse were used as response keys. Participants responded with the index fingers of both hands. Participants sat 60cm from the screen in a quiet, dimly lit lab.

2.2.   Experiment 1 - results

Thirteen participants were discarded from our analyses because of low accuracy in the sentence recognition task. The remaining 41 participants had accuracies > 90% in both tasks. After the experiment, participants were asked whether they ever noticed a number (the masked P_2) appearing during the presentation of the two random strings of characters. Nobody reported noticing P_2.

In analysing number evaluation trials, we considered only responses that were 250ms < RT < 1000ms (92.5% of all responses). This procedure is consistent with existing priming research (e.g, Kouider & Dehaene, 2009; Naccache & Dehaene, 2001a, 2001b). We adopted it to remove possible outliers (i.e., either anticipatory or delayed responses) and to keep our findings comparable with those already available in literature. As a first step, we assessed the presence of priming effects in the Modus Ponens ("if p then q; p; q") and Disjunctive Syllogism ("not p or q; p; q"; "p or not q; q; p") problems by comparing the RTs of the critical problems with the RTs of the relevant baselines

(which we refer to as ΔRT). Then, we tested for priming effects in ΔRT in the other relevant problems to evaluate whether the effects possibly observed for Modus Ponens or Disjunctive Syllogism were specific to the type of problem. The additional control conditions involved: (*i*) all other problems that validly or invalidly entailed T (e.g., Affirmation of the Consequent) and (*ii*) all problems that did not support T, but bore a strong similarity to Modus Ponens or Disjunctive Syllogism problems, differing from them in one feature only (e.g., Modus Ponens *vs.* "p or q; p; q"; Modus Ponens *vs.* "if p then not q; p; q").

A preliminary ANOVA factoring the two critical problems (Modus Ponens and Disjunctive Syllogism) and the ISI condition (long and short) with ΔRT as the dependent variable showed that the interaction was not significant ($F(1,39) = .35$; $p = .55$). Data from the short and long ISI conditions were therefore pooled together.

Next, we tested for priming effects in Modus Ponens and Disjunctive Syllogism problems. We found a priming effect for Modus Ponens (ΔRT = -20.6, $\eta_p^2 = .162$, $t(40) = 2.77$, $p = .008$) but not for Disjunctive Syllogisms (ΔRT = 9.7, $\eta_p^2 = .059$, $t(40) = 1.58$, $p = .121$). Subjects were *faster* at answering "odd" or "even" when the target represented the conclusion of a Modus Ponens inference. Given this priming effect for Modus Ponens, we ran planned control analyses. The comparisons and results are detailed in Table 2. Overall, no priming effect was observable in any of the control problems suggesting that the effect we found for Modus Ponens is highly specific to the type of problem. To further test this conclusion, we directly compared the ΔRT of Modus Ponens with the ΔRT of all control problems. The ΔRT was significantly smaller for Modus Ponens than for all other control trials (ΔRT = -21 *vs.* -1; $t(40) = 2.31$, $p = .026$). The ΔRT of the

Modus Ponens was also smaller than the mean ΔRT related to all other types of trials (ΔRT = -21 *vs.* 2; $t(40) = 3.06$, $p = .004$). Furthermore, we compared the RT on Modus Ponens with a more "general" baseline, consisting in the average RT over all types of conditional baseline trials (i.e., any conditional trial using "r" as P_2). The RT on Modus Ponens was faster than the general baseline ($t(40) = 3.94$, $p < .001$). By contrast, none of the other control problems differed from the general baseline (all *ps* > .1).

Table 2. Test of priming effects on the Modus Ponens inference and the relevant control conditions.

P_1	P_2	T	RT exp trial (SD)	ΔRT (SD)	$t(40)$	p	η_p^2	Description of the experimental trial
If p then q	p	q	601 (100)	- 21(47)	2.78	.008	.162	Target matches Modus Ponens inference
If p then q	p	p	621 (104)	- 11 (71)	0.96	.34	.022	Target does not match Modus Ponens inference
If p then not q	p	q	619 (102)	2 (54)	0.26	.79	.002	Target does not match Modus Ponens inference
If not p then q	p	q	617 (103)	-12 (61)	1.24	.22	.037	Target does not match any inference from the premises
If not p then not q	q	p	635 (101)	6 (59)	0.63	.53	.010	Target matches Modus Tollens inference
If p then q	q	p	622 (100)	-10 (60)	1.02	.32	.025	Target matches Affirmation of the Consequent fallacious inference
If not p then not q	p	q	630 (106)	9 (54)	1.13	.27	.031	Target matches Denial of the Antecedent fallacious inference
p or q	p	q	626 (105)	16 (57)	1.77	.08	.073	Target matches a Disjunctive Fallacy

ΔRT: Difference in RT between the experimental problems and the relevant baseline problems.

Finally, no priming effects for Modus Ponens or Disjunctive Syllogisms were found when separate analyses were carried out on the data from the 13 participants who were discarded due to low performance in the sentence recognition task. This finding suggests that an adequate representation of the first premise in WM is a necessary condition for the observation of the priming effect in Modus Ponens trials.

2.3. Experiment 1 - discussion

The findings of Experiment 1 are interesting in several respects. First, we showed that judging parity is faster in trials in which the number to be evaluated is the conclusion of a single step Modus Ponens inference ("if p then q; p; q") than in baseline trials, that were identical except for the non-matching second premise ("if p then q; *r*; q"). Critically, this was the case even though, in our paradigm, the second premise was reportedly *invisible* to the subjects. Thus, as far as the subjective experience of the participants was concerned, the Modus Ponens problems and the relevant baseline problems were indistinguishable.

Second, the observed advantage in RTs was *only* found for the Modus Ponens problems (Table 2). The effect was not present in the Affirmation of the Consequent fallacy ("if p then q; q; p"), thus showing that the effect was not based on a bidirectional association of the antecedent ("p") and consequent ("q") of the conditional premise. Furthermore, the effect was not present in trials with a sequence of numbers *identical* to Modus Ponens ("p.q.p.q") but containing negations (e.g., "if not p then q; p; q") or containing a disjunction instead of a conditional connective ("p or q; p; q"). Thus, the effect observed for Modus Ponens cannot be merely considered an association of the two arguments (p and

q) of the conditional rule independent of the presence of negations and the type of logical connective involved.

Third, trials in which Modus Ponens would be triggered by an indirect match between P_2 and P_1 did not show the effect. This suggests that the match between P_2 and the antecedent in P_1 needs to be straightforward (e.g., an identity between the antecedent in P_1 and P_2) to automatically trigger the Modus Ponens inference.

Fourth, the Disjunctive Syllogisms ("not p or q; p; q"; "p or not q; q; p") did not demonstrate a similar priming effect relative to their baseline trials ("not p or q; r; q"; "p or not q; r; p"). Similar to preceding studies arguing for the automaticity of some basic inferential schemata (Lea, 1995), we expected that both the Modus Ponens and the Disjunctive Syllogism would be triggered automatically by an unconscious premise or that neither would. Instead, the results strongly suggest that Modus Ponens is a "special" type of inference, one that can be triggered even by invisible premises, seemingly in the absence of conscious willingness to reason, whereas the Disjunctive Syllogism – apparently – is not drawn if the minor premise is not seen.

3.   Experiment 2: visibility check

No participant in Experiment 1 reported noticing the presentation of any number after P_1. They were aware only of having seen a string of flickering letters (the masks). Even though this *subjective* measure is valid for establishing that P_2 was subjectively non-conscious according to some authors and theories (e.g., Seth, Dienes, Cleeremans, Overgaard, & Pessoa, 2008), according to others,

it is not exhaustive (e.g., Kouider & Dupoux, 2001). Thus, the aim of this control experiment was twofold. First, we wanted to replicate the finding that, in the same experimental conditions as in Experiment 1, participants do not report noticing P_2 when they are explicitly questioned about it after the experiment. Second, and more importantly, we wanted to check whether participants who were alerted to the presence of P_2 between the two masks could access information about it.

3.1. Methods

3.1.1. Participants

Twenty-three (9 females) graduate and undergraduate students from the University of Milano-Bicocca participated in the experiment in exchange for course credit.

3.1.2. Stimuli

Experiment 2 was composed of two independent tasks, administered in the same order.

3.1.2.1. Task 1. The first task was a shortened version of Experiment 1. The stimuli were the same, but only 128 *number evaluation trials* (i.e., two trials per problem type) and 32 *sentence recognition trials* were used. After that, a *debriefing questionnaire* was administered, asking whether the participants saw anything between the two strings of random characters, and – if so – what they saw.

3.1.2.2. Task 2. Task 2 was administered immediately after Task 1. The same stimuli as in Experiment 1 were used, with the exception that half the time P_2

was a number and half the time a letter. Subjects were alerted to the presence of a number or a letter between the two strings of random characters and were explicitly asked to both *(i)* read and retain P_1 in memory and *(ii)* try as hard as possible to detect whether P_2 was a number or a letter. Sixty-four *sentence recognition trials*, identical to those in Experiment 1, were administered to assure that the WM load in Task 2 was similar to that in Experiment 1. Additionally, 256 *detection trials* were administered, 128 displaying a number in P_2 and 128 displaying a letter. These trials were a substitute for the number evaluation trials in Experiment 1. In detection trials, the question "Was there a number or a letter between the two strings of characters?" was displayed on the screen after the second mask. Participants were told to choose a category, even if they were uncertain or had the impression of having seen nothing. Sentence recognition trials and detection trials differed only during question/target presentation. Participants could know whether it was a sentence recognition trial or a detection trial only at the end of a trial. Thus, the processing of both kinds of trials was identical until the final phase. Detection trials and sentence recognition trials were randomly interspersed.

3.1.3. Procedure

The timeline for both tasks was the same as in Experiment 1, except that the ISI after the second mask was always set at 0ms. There were no time limits for responding to detection trials. The first task was preceded by a training session of 40 trials with feedback on accuracy. The main dependent variable of the first task was the response to the debriefing questionnaire. A 10-trial training session with accuracy feedback preceded the second task. The main dependent variable

of the second task was the *d'* (sensitivity index, i.e., a measure comparing the *hit rate* and the *false-alarm rate*) in the detection trials. In the current study, we lowered the recognition-accuracy threshold in the sentence recognition trials from 90% to 80% accuracy because participants performed worse on this measure than in the previous study due to the high cognitive demand required to detect P_2 (using a 90% threshold would exclude 61% of participants). This choice implies that participants in this task had – on average – a slightly reduced WM load compared to those in Experiment 1. Nevertheless, please notice that a lower exclusion threshold should, if anything, decrease priming effects.

3.1.4. Apparatus

The experiment was programmed and administered using Presentation[R] (Neurobehavioral Systems, Inc., San Francisco, CA, US). The CRT screen had a 60Hz refresh rate and the resolution was set at 1024 × 768 pixels. The premises, masks, and target stimuli were written in white, 22-point Arial characters and were presented against a black background. Two buttons of a serial mouse were used as response keys. Participants responded with the index fingers of both hands. Participants sat 60cm from the screen in a quiet, dimly lit lab.

3.2. Experiment 2 - results

3.2.1. Task 1

None of the participants reported having seen a number (P2) between the masks. Thus, the task confirmed that participants did not notice P2, when they were not alerted to the presence of P2 and were not explicitly told to detect it.

Notwithstanding the low power of Experiment 2, we also checked whether the main findings of Experiment 1 replicated in Experiment 2. Seven participants

had to be excluded because of accuracy scores < 80% in the sentence recognition task and three subjects had to be excluded because they had less than two trials per condition after removal of incorrect trials. This left 13 participants for the analyses. Given that Experiment 1 established the directionality of the priming effect for Modus Ponens, we used one-tailed statistical tests. A priming effect specific for Modus Ponens was found also in Experiment 2. The Modus Ponens trials were faster compared to the relevant baseline ("If p then q; p; q" *vs.* "If p then q; r; q"; $t(12) = 2.43$, $\eta_p^2 = .33$, $p = .016$). By contrast, the priming effect was not present in the Affirmation of the Consequent fallacy ("If p then q; q; p" *vs.* " If p then q; r; p"; $t(12) = 1.09$, $\eta_p^2 = .09$, $p = .15$) and the Disjunctive problem similar to Modus Ponens ("p or q; p; q" *vs.* "p or q; r; q"; $t(12) = .13$, $\eta_p^2 = .001$, $p = .45$). As in Experiment 1, we directly compared the ΔRT of Modus Ponens with the ΔRT of all control problems. The ΔRT was significantly smaller for Modus Ponens than for all other control trials ($t(12) = 2.32$, $p = .019$). The ΔRT of the Modus Ponens was also smaller than the mean ΔRT related to all other types of trials ($t(12) = 2.60$, $p = .012$).

3.2.2. Task 2

Nine participants were excluded from the following analyses because of accuracy scores < 80% in the sentence recognition task. Of the remaining 14 participants, one was strongly biased toward the "letter" response: she was 100% correct when P_2 was a letter (not allowing us to compute her d'), but 80% of the time confused numbers for letters. The mean d' of the remaining 13 participants was .46 (median = .37), which is significantly greater than zero ($t(12) = 4.58$, $p = .001$). This finding corresponds to an overall accuracy rate of

56% with 11% more hits than false alarms, meaning that – in aggregate form – they were moderately able to discriminate whether the second premise was a letter or a number.

3.3. Esperiment 2 - discussion

The first task used a subjective measure of consciousness (Seth et al., 2008) and confirmed that in an experimental setting similar to Experiment 1 people did not report having seen P_2. Furthermore, Experiment 2 replicated the main findings of Experiment 1. Namely, a priming effect was found for Modus Ponens ("if p then q; p; q") but not for the Affirmation of the Consequent ("if p then q; q; p") and for the disjunctive trial similar to Modus Ponens ("p or q; p; q"). Two facts rule out the possibility that the effect we observed was learned during our experimental procedure: 1) The effect could be found even after the administration of only two Modus Ponens trials with the target matching the conclusion and 2) No feedback was provided during Experiment 1 or 2.

In the second task, participants were informed of the possible presence of a number in P_2 and were asked to keep track of it. This instruction produced an expected reallocation of attention to P_2, as indexed by a poorer overall performance in the sentence recognition trials. In these more favourable conditions, the subjects were able to judge whether P_2 was a letter or a number with better-than-chance accuracy. It must be noted, however, that – even though alerted to the presence of P_2 – participants attained a rather low d'. Overall, our findings suggest that the P_2 premises in Experiment 1 were probably "preconsciously" processed, which, according to a distinction proposed by Dehaene and collaborators, means that they were only weakly processed but

could be consciously detected if enough attention were devoted to them (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; Kouider & Dehaene, 2007). Accordingly, it is possible that, in a few trials, some participants in Experiment 1 were conscious of P_2 even though they might have considered the experience too weak to be reported. It should be noted that this result does not change the fact that parity judgments in Experiments 1 and 2 were selectively enhanced when the target numbers matched the valid conclusion of a Modus Ponens inference. However, the findings in Experiment 2 weaken the claim that a Modus Ponens inference can be triggered even if its minor premise is presented *subliminally*. It is possible that the advantage on Modus Ponens in Experiments 1 and 2 depends on the few trials in which the participants might have accessed some information about P_2.

4. Experiment 3

Experiment 3 had two goals. First and foremost – following Experiment 2 – we wanted to check whether the priming effect observed for Modus Ponens in Experiment 1 was present in subjects for whom processing of P_2 could be considered *subliminal*. Thus, in Experiment 3, we considered only those subjects who showed chance performance in identifying P_2 in a forced-choice detection task. Furthermore, the procedure adopted in Experiment 3 for assessing conscious access to P_2 improved compared with Experiment 2. In Experiment 2 we only checked whether the participants could access relatively superficial information, i.e., whether a stimulus was a number or a letter. However, it is possible that the participants had conscious access only to *some* information

about P_2 (e.g., "number or letter?") but not to other (e.g., "which number was presented?"). Given that the critical information to trigger Modus Ponens or other inferences is indeed the identity of P_2, in Experiment 3 we tested more specifically for the availability of this information (see also Naccache & Dehaene, 2001b for a similar line of reasoning).

The second goal was to test whether the advantage for Modus Ponens could also be observed when the conditional major premise P_1 was presented in an inverted form, namely with the logical consequent (in this case p) preceding – linguistically – the logical antecedent (q). To accomplish this, we added a subset of "inverted Modus Ponens" trials, whose form was "p if q; q; p" (e.g., P_1: "There is an 8 if there is a 3"; P_2: "3"; T: "8"), and the corresponding baseline trials ("p if q; r; p") to the main task.

4.1. Methods

4.1.1. Participants

Forty-three (35 females) graduate and undergraduate students from the University of Milano-Bicocca participated in the experiment in exchange for course credit.

4.1.2. Stimuli

4.1.2.1. Task 1. Task 1 consisted of a shorter version of Experiment 1. Specifically, for the *number evaluation trials*, negations of P_1 in the experimental trials and some of the original filler trials were not used, leaving only four experimental problems and their corresponding baselines. To these, we added two experimental problems with an inverted conditional in P_1 of the form "p if q" and their baselines. To avoid always presenting target numbers that were also

mentioned in P_1, we introduced six filler conditions in which a different number was used as the target. The full list of the types of trials in the number evaluation task is reported in Table 3. The numbers used as p, q, or r were randomly chosen from numbers between 1 and 8.

Finally, to reduce the similarity of the visible sentences in the task, we also included some trials where we randomly inserted negations of P_1. Overall, we administered 32 replications for each experimental cell and 96 trials with random negations, corresponding to $32 \times 18 + 96 = 672$ trials in the number evaluation condition. We also administered 130 *sentence recognition trials*, in which the features of P_1 (type of logical connective and presence or absence of negations) and P_2 (the numbers p, q, or r) were independently and randomly manipulated.

Table 3. Types of problems used in Experiment 3.

Type	Problem	Baseline	Filler
Standard Modus Ponens	If p then q; p; q	If p then q; r; q	If p then q; p; r
Standard Affirmation of the Consequent	If p then q; q; p	If p then q; r; p	If p then q; q; r
Disjunctive fallacy 1	p or q; p; q	p or q; r; q	p or q; p; r
Disjunctive fallacy 2	p or q; q; p	p or q; r; p	p or q; q; r
Inverted Modus Ponens	p if q; q; p	p if q; r; p	p if q; q; r
Inverted Affirmation of the Consequent	p if q; p; q	p if q; r; q	p if q; p; r

4.1.2.2. Task 2. Task 2 was administered immediately after Task 1. In Task 2, we used a set of stimuli similar to Task 1, with the exception that P_2 was always a random number between 0 and 9. Subjects were alerted to the presence of a number between the two strings of random characters and were *explicitly* asked: (*i)* to read and retain P_1 in memory and (*ii)* to try as hard as possible to identify P_2. Ninety-six *detection trials* were administered. In addition, 48 *sentence*

recognition trials, identical to those in Task 1, were presented to make sure that the WM load in this task was the same as in Task 1. In *detection trials*, the question "Which number was there between the two strings of random characters?" was displayed on the screen after the second mask. Participants were told to choose a number and press the corresponding key on the PC keyboard, even in trials in which they were highly uncertain about the identity of P_2. Sentence recognition trials and detection trials differed only during question/target presentation. Participants could know whether it was a sentence recognition trial or a detection trial only at the end of a trial. Thus the processing of both kinds of trials was identical until the final phase. Detection trials and sentence recognition trials were randomly interspersed. There were no number evaluation trials in Task 2.

4.1.3. Procedure

The timeline of both tasks is the same as in Experiment 1, except that the ISI after the second mask varied across four levels: 0, 250, 500 or 750ms. In Task 2, subjects had no time limit to respond to detection trials. The first task was preceded by a training session of 50 trials with feedback on accuracy. The second task was preceded by a 10-trial training session with accuracy feedback.

4.1.4. Apparatus

The experiment was programmed and administered using Presentation[R] (Neurobehavioral Systems, Inc., San Francisco, CA, US). The CRT screen had a 60Hz refresh rate and the resolution was set at 1024 × 768 pixels. The premises, masks, and target stimuli were written in white, 22-point Arial characters and were presented against a black background. Two buttons on the Cedrus™

Response Box RB-730 were used as response keys in Task 1. Participants responded with the index fingers of both hands. In Task 2, responses were collected via the keyboard. Participants sat 60cm from the screen in a dimly lit lab. To limit distraction due to ambient phasic noise, participants wore Sennheiser HD280 headsets, providing 32db of attenuation in external noise. Finally, white noise was played in the headset to further isolate the subject. A photodiode was connected to an oscilloscope to check exactly how long the short P_2 stimuli were displayed on the screen. The oscilloscope confirmed that the stimuli were displayed for, at most, 50ms as requested. More specifically, as expected on CRT monitors, the brightness of the stimuli started to decay slightly before 50ms so that at 40ms the brightness was about 1lux, at 45ms about 0.5lux, and at 50ms less than 0.1lux.

4.2. Experiment 3 - results

4.2.1. Task 2

Average accuracy on the sentence recognition trials in Task 2 was 91% and no participant responded at chance. The average accuracy on the detection task was 21%. This was greater than the accuracy level expected by chance alone (10%, $t(42) = 5.19$, $p < .001$) showing that, on average, subjects were able to get some information about the identity of P_2, even though performance was far from optimal, closely replicating the results of Experiment 2. The correlation between accuracy on the sentence recognition task and on the detection task was negligible (r = - .11, $p > .1$); this assures that the differences in performance in the detection task were not due to a general disengagement from the task. The ability to identify P_2 varied widely across subjects, from chance-level to a

maximum of 61% accuracy. In particular, a subset of 18 subjects performed at chance (< 16% accuracy) as determined by a binomial test. The average accuracy of these subjects was 10.7% (Standard Deviation, SD = 3%) and was not significantly different from chance ($t(17) = .99$, $p > .1$).

4.2.2. Task 1

The main objective in Experiment 3 was to check whether the priming effect observed in the Modus Ponens trials could be replicated for participants who did not have conscious access to the identity of P_2. Accordingly, we applied our main analyses to the 18 subjects with chance performance on Task 2. Given that Experiment 1 established the directionality of the priming effect for Modus Ponens, we used one-tailed statistical tests. As in Experiment 1, in analysing number evaluation trials we considered only responses that were 250ms < RT < 1000ms. The pattern of results closely replicated the main findings of Experiment 1. In all trials and conditions, subjects were accurate and fast in judging whether the target number presented was odd or even: average accuracy was 96% (SD = 3%) and the average RT was 660ms (SD = 108ms). Most importantly, however, subjects were faster in judging whether a target number was odd or even when the target was the conclusion of a Modus Ponens: trials "if p then q; p; q" were faster than trials "if p then q; r; q" (ΔRT = -14.0ms, η_p^2 = .242, $t(17) = 2.33$, $p = .016$). As in Experiment 1, this advantage was specific to the Modus Ponens problems; the effect was not present in any of the other conditions. In particular, we did not find any advantage for the Affirmation of the Consequent fallacy "if p then q; q; p" over the corresponding baseline "if p then q; r; p" (ΔRT = -4.2ms, η_p^2 = .014, $t(17) = .484$, $p > .1$) and for the relevant

disjunctive fallacy "p or q; p; q" *vs.* "p or q; r; q" (ΔRT = 1.0ms, $\eta_p{}^2$ = .001, t(17) = .11, p > .1). To further test this conclusion, we directly compared the ΔRT of Modus Ponens with the ΔRT of all other problems, but inverted Modus Ponens. The ΔRT of the Modus Ponens showed a robust statistical trend in being smaller than the mean ΔRT related to all other types of trials (ΔRT = -14.0 *vs.* -4.3; t(17) = 1.67, p = .056). The inverted Modus Ponens "p if q; q; p" trials did not reach a significant ΔRT with respect to their baseline, but they showed a trend in the expected direction (ΔRT = -10.2, $\eta_p{}^2$ = .126 t(17) = 1.56, p = .068). Conversely, the inverted Affirmation of the Consequent "p if q; p; q" showed no advantage over its control baseline (ΔRT = -6.4ms, $\eta_p{}^2$ = .037, t(17) = .81, p > .1).

Subjects in this subgroup had an average accuracy on the sentence recognition task of 88% (SD = 6%). All subjects had an accuracy level higher than chance. Given the relatively small size of the subgroup, we could not apply a selection criterion based on the sentence recognition performance as stringent as in Experiment 1, as it would have left only eight subjects for analysis. However, we applied a more moderate threshold (performance on sentence recognition trials higher than 80%) and found that the remaining 16 subjects showed the same pattern of performance as the whole sample, but more pronounced. The Modus Ponens condition was faster than baseline (ΔRT =-17.1ms, $\eta_p{}^2$ = .437, t(15) = 3.41, p = .002), while the other control conditions were not (Affirmation of the Consequent: ΔRT = -5.4ms, $\eta_p{}^2$ = .023, t(15) = .59, p > .1; disjunctive fallacy: ΔRT = 1.0ms, $\eta_p{}^2$ = .001, t(15) = .12, p > .1). Finally, the inverted Modus Ponens condition still showed a statistical trend towards being faster than the baseline (ΔRT =-10.3ms, $\eta_p{}^2$ = .133, t(15) = 1.51, p = .075),

whereas the inverted Affirmation of the Consequent condition did not show any advantage (ΔRT =-5.1ms, η_p^2 = .032, t(15) = .70, p > .1).

4.3. Experiment 4 - discussion

Experiment 3 replicated and extended the main finding of Experiment 1, namely the positive priming effect caused by the automatic triggering of a Modus Ponens inference. We tested whether the priming effect for Modus Ponens observed in Experiment 1 could be found also in subjects who processed P_2 *subliminally*. To do this, we selected participants that were at chance in a forced-choice detection task on P_2. Even though these subjects did not have either conscious or pre-conscious access to P_2, the priming effect for the standard Modus Ponens was still observed. Furthermore, as in Experiment 1, this effect was specific: it was absent in other highly similar conditions, namely the Affirmation of the Consequent ("if then q; q; p") and the disjunctive problem with a sequence of arguments paralleling the standard Modus Ponens argument ("p or q; p; q").

A further objective of Experiment 3 was to test whether the effect we found for Modus Ponens inferences in Experiment 1 would generalise to Modus Ponens inferences relying on an alternative, non-standard representation of the problem. We considered an "inverted" version of the Modus Ponens in which the logical consequent linguistically preceded the logical antecedent (i.e., the trials in which P_1 had the form "there is p if there is q"). In addition to Modus Ponens, we also considered an "inverted" version of the Affirmation of the Consequent fallacy. Notice that the sequence of literals in the inverted fallacy is identical to the sequence of literals in the standard Modus Ponens ("p.q.p.q") and,

analogously, the sequence of literals in the inverted Modus Ponens is identical to the standard fallacy. The findings were in the expected direction. The inverted fallacy did not show any effect, while the inverted Modus Ponens showed a statistical trend. The fact that the inverted fallacy was not significant corroborates evidence from Experiment 1 showing that the mere consideration of the sequence of the literals ("p.q.p.q") is *not* enough to justify the observed effect for Modus Ponens. Nevertheless, the weaker effect shown for the inverted Modus Ponens suggests that the two alternative representations of P_1 are not equivalent (see also Grosset & Barrouillet, 2003).

5.   Experiment 4

In Experiments 1-3 we showed that subjects judged faster whether a digit was odd or even when it was the conclusion of a Modus Ponens. We also showed that this effect was present *specifically* in Modus Ponens but not in a wide set of other conditional and disjunctive problems. In Modus Ponens, the conditional statement ("if p then q") instantiates a unidirectional linkage from the antecedent "p" to the consequent "q". Thus, it may be possible that other kinds of relations able to produce a "unidirectional linkage" between antecedent and consequent would generate a priming effect, as we observed for Modus Ponens. In Experiment 4, we aimed to explore whether the priming effect observed for Modus Ponens was present in problems instantiating other kinds of relations inducing a unidirectional linkage between the two terms. In particular, we considered spatial relations (e.g., "*a* before *b*") and sentences with quantifiers (e.g., "all *a* with *b*"). Besides these new conditions, we also administered the

Modus Ponens problem to check for the presence of the effect in another subject sample.

5.1. Methods

5.1.1. Participants

Twenty-seven (18 females) graduate and undergraduate students from the University of Milano-Bicocca participated in the experiment in exchange for course credit.

5.1.2. Stimuli

The administered task was a modified version of Experiment 1. For the *number evaluation trials*, we used eight experimental conditions and their corresponding baselines (Table 4). Two of the eight experimental problems were a replication of two conditions already administered in all preceding experiments: the Modus Ponens and the Affirmation of the Consequent. The other six experimental problems were new conditions exploring the possible priming effects related to *spatial relations* (e.g., "There is a 2 before the 4") and *quantified propositions* (e.g., "Every 2 is together with a 4") that suggested a directional association between the two terms. To avoid always presenting target numbers that were also mentioned in P_1, we introduced eight filler conditions in which a different number was used as a target (Table 4). Overall, we administered 16 replications for each experimental condition, corresponding to 384 trials in the number evaluation condition. We also administered 192 *sentence recognition trials*, in which the features of P_1 (type of logical connective and presence or absence of negations) and P_2 (the numbers p, q, or r) were independently and randomly manipulated. Thus, overall, 576 experimental trials

were administered. A *debriefing questionnaire* was administered after completion of the main task, exploring whether the participants saw digits between the two strings of random characters and – if so – how many times.

Table 4. Types of problems used in Experiment 4.

Type	Problem	Baseline	Filler
Modus Ponens	If p then q; p; q	If p then q; r; q	If p then q; p; r
Affirmation of the Consequent	If p then q; q; p	If p then q; r; p	If p then q; q; r
Spatial relation 1	p before q; p; q	p before q; r; q	p before q; p; r
Spatial relation 2	p before q; q; p	p before q; r; p	p before q; q; r
Quantified relation 1	Every p with q; p; q	Every p with q; r; q	Every p with q; p; r
Quantified relation 2	Every p with q; q; p	Every p with q; r; p	Every p with q; q; r
Quantified relation 3	No p with q; p; q	No p with q; r; q	No p with q; p; r
Quantified relation 4	No p with q; q; p	No p with q; r; p	No p with q; q; r

5.1.3. Procedure

The timeline of the task was the same as in Experiment 1, except that the ISI between the second mask and the target was always set at 0ms. The task was preceded by a training session of 50 trials with feedback on accuracy.

5.1.4. Apparatus

The experiment was programmed and administered using Presentation[R] (Neurobehavioral Systems, Inc., San Francisco, CA, US). The CRT screen had a 60Hz refresh rate and the resolution was set at 1024 × 768 pixels. The premises, masks, and target stimuli were written in white, 22-point Arial characters and were presented against a black background. Two buttons on the Cedrus[TM] Response Box RB-730 were used as response keys. Participants responded with the index fingers of both hands. To limit distraction due to ambient phasic noise, participants wore Sennheiser HD280 headsets, providing 32db of attenuation in

external noise. Finally, white noise was played in the headset to further isolate the subject.

5.2.   Experiment 4 - results

Average accuracy in the sentence recognition trials was 96% (SD = 3%). All subjects had an accuracy level above 90% but one, who had an accuracy level between 80% and 90%. Subjects were accurate in judging whether the target numbers presented were odd or even: average accuracy was 95% (SD = 4%). All subjects had an average accuracy above 90% but one, who had an average accuracy between 80% and 90%. The subjects were fast at answering the number evaluation trials: the average RT on correct trials was 635ms (SD = 106ms). As in Experiment 1, in analysing number evaluation trials we considered only responses that were 250ms < RT < 1000ms and subject who had a high accuracy (> 90%). When specifically asked, four subjects reported having noticed the presence of digits between the two strings of random characters *less* than ten times out of 576 trials. All the other subjects did not notice that a digit was presented.

As in Experiments 1-3, subjects were faster in judging whether a target number was odd or even when it was the conclusion of a Modus Ponens: The subjects were faster in trials "if p then q; p; q" than in "if p then q; r; q" (ΔRT = -16.9ms, $\eta_p{}^2$ = .125, $t(24)$ = 1.86, p = .035). This advantage was specific to the Modus Ponens problems: The effect was not present in any of the other conditions. In particular, we did not find any advantage for the Affirmation of the Consequent fallacy "if p then q; q; p" over the corresponding baseline (ΔRT = 4.70ms, $\eta_p{}^2$ = .016, $t(24)$ = -.625, p > .1). This was also the case for the two spatial

relations, "p before q; p; q" (ΔRT = 10.31ms, $\eta_p{}^2$ = .078, t(24) = -1.428, p > .1) and

"p before q; q; p" (ΔRT = 4.07ms, $\eta_p{}^2$ = .012, t(24) = -.539, p > .1), and for all the

four quantified relations: "Every p with q; p; q" (ΔRT = -3.13ms, $\eta_p{}^2$ = .009, t(24)

= 0.464, p > .1); "Every p with q; q; p" (ΔRT = 14.4ms, $\eta_p{}^2$ = .12, t(24) = -1.82, p >

.1); "No p with q; p; q" (ΔRT = 11.94ms, $\eta_p{}^2$ = .079, t(24) = -1.43, p > .1); "No p

with q; q; p" (ΔRT = -6.60ms, $\eta_p{}^2$ = .016, t(24) = .62, p > .1). To further test this

conclusion, we directly compared the ΔRT of Modus Ponens with the ΔRT of all

other problems. The ΔRT of the Modus Ponens was also smaller than the mean

ΔRT related to all other types of trials (ΔRT = -18.7 *vs.* 2.8; t(26) = 2.27, p = .016).

The overall pattern of performance did not change when we excluded from the

analyses the four subjects who noticed (less than 10 times) the presence of P$_2$

(Modus Ponens: ΔRT = -21.4ms, $\eta_p{}^2$ = .176, t(20) = 2.069, p = .026). The pattern

did not change also when the two participants who had an accuracy level

between 80% and 90% in one of the two tasks (sentence recognition and

number evaluation) were considered in the analyses as in Experiment 2 and

Experiment 3. (Modus Ponens: ΔRT = -18.7ms, $\eta_p{}^2$ = .149, t(26) = 2.13, p = .021).

5.3. Experiment 4 - discussion

The findings of Experiment 4 confirmed the existence and the specificity of

the priming effect in the Modus Ponens problems.

First, Experiment 4 replicated the main findings of Experiments 1-3: People

were faster in evaluating the parity of a number when the number was the

conclusion of a Modus Ponens inference, even though they were not required to

make any inference and they were not subjectively aware that the second

premise was presented.

Second, we further confirmed that the priming effect for Modus Ponens is highly specific. The priming effect, as in Experiments 1-3, was absent in the Affirmation of the Consequent problems. More interestingly, the priming effect was also absent in problems exploring spatial and quantified relations, suggesting that the observed effect is specific to conditional sentences. Furthermore, it should be noticed that a subset of the spatial and quantified problems shared with Modus Ponens the same sequence of presented digits (compare "If p then q; p; q" with "p before q; p; q" or "Every p with q; p; q" or "No p with q; p; q"). The absence of a priming effect in these cases corroborates the conclusion that presenting a sequence of numbers similar to those involved in Modus Ponens ("p.q.p.q") is *not* a sufficient condition to induce a faster processing of the target number. Finally, it should be pointed out that *neither* is logical validity. This had already been shown in Experiment 1 with the Modus Tollens and the Disjunctive Syllogism. Experiment 4 added a further valid problem to the list: "Every p with q; p; q".

Overall, the effect observed for Modus Ponens problems seems to depend on the specific representation induced by the conditional sentence "If p then q".

6. General discussion

In the present study, we investigated whether two basic deductive inferences at the core of human rational thought (Rips, 1988, 1994), the Modus Ponens and the Disjunctive Syllogism, are automatically and unconsciously processed. Furthermore, we investigated whether these inferences require a fully explicit representation of both premises or they can also be triggered by

minor premises that are not consciously perceived. In Experiment 1, the major premises (P_1) of various valid or invalid inferential schemata were explicitly displayed and retained in memory, but the corresponding minor premises (P_2) were very rapidly displayed and masked (see also Naccache & Dehaene, 2001b). Participants were not alerted to the presence of a minor premise or to the possibility that deductive inferences could be drawn during the task. In a post-experimental debriefing session, subjects reported that they did not notice any P_2. Thus, according to their self reports, two trials sharing the same first premise (for instance, "if there is a 3 then there is an 8") and the same target (for instance, "8") but with different second premises (for instance, "3" and "5") were perceived as identical trials, since P_2 was not noticed. Surprisingly, when the unnoticed second premise corresponded to the antecedent of the conditional ("3" in the example above), participants pre-activated the number "8", as shown by a positive priming effect on an ensuing number evaluation task displaying the same number ("8"). This finding suggests that Modus Ponens inferences are not only automatic but can be carried out unconsciously and can be triggered by unnoticed premises. Intriguingly, the Modus Ponens was the only schema that showed these properties, when triggered by a direct match on the rule's antecedent. No priming effects were observed neither for the Disjunctive Syllogism nor for any other set of premises allowing valid or invalid inferences (e.g., Affirmation of the Consequent, Denial of the Antecedent, disjunctive fallacies, Modus Tollens). Thus, we suggest that the observed priming effect did not depend exclusively on non-logical associations between the arguments of the major premise, but on some peculiarity in the way the brain deals with the Modus Ponens logical structure.

Findings from Experiment 1 were replicated and extended in three follow-up experiments. First, we checked whether the participants were, indeed, unable to access the minor premises as they reported in Experiment 1. Experiment 2 suggested that some information about P_2 was actually available to some participants, at least when their attention had been directed to P_2 by telling them about the masked prime. Given the results of Experiment 2, it was not possible to fully establish whether the effect on Modus Ponens required *pre-conscious* access to P_2 or, alternatively, whether it would also have been present in the case of a *subliminal* processing of P_2 (Dehaene et al., 2006). Accordingly, we ran a third Experiment in which we focused on the subset of participants for whom subliminal processing of P_2 had been demonstrated. The priming effect for Modus Ponens was also confirmed in Experiment 3. Finally, in Experiment 4, we explored whether relations other than those involved in propositional logic would produce a priming effect similar to the one observed for Modus Ponens. In particular, we explored spatial relations (e.g., "*a* before *b*") and sentences with quantifiers (e.g., "all *a* with *b*"). By contrast with the priming effect for Modus Ponens, which was also replicated in Experiment 4, we could not observe a priming effect for any of the other types of relations.

These findings shed new light on our understanding of human deduction. First, they suggest that Disjunctive Syllogism is not automatic, or is less automatic than the Modus Ponens, a conclusion that is partially at odds with the predictions made according to Mental Logic Theory (Braine & O'Brien, 1998). This theory proposes that both inferences are part of a direct routine of reasoning, each requiring the activation of only one direct reasoning schema. Consequently, the theory correctly predicts the automaticity of the Modus

Ponens, but also the theory wrongly predicts the automaticity of the Disjunctive Syllogism.

Second, the results suggest that subliminal processing of conditional premises is unidirectional, going from the antecedent to the consequent. The Affirmation of the Consequent fallacy (i.e., "If p then q; q; therefore p") is not automatic, or is less automatic than the Modus Ponens. This is at odds with the predictions made by the Mental Model Theory (Johnson-Laird & Byrne, 1991; Rader & Sloutsky, 2002), which assumes that the initial understanding of a conditional rule consists in the mental representation of a possibility, in which the antecedent and consequent coexist, with no directionality assumptions. If confirmed, these results would prompt a revision of those influential cognitive theories of human deductive reasoning.

Third, many theories have suggested that a unique reasoning machinery deals with all basic deductive inferences in essentially the same way (Braine & O'Brien, 1998; Johnson-Laird, 2006). Instead, it seems that one important and very common inference, the Modus Ponens, behaves in a *qualitatively different* way from all the other inferences we tested. Contrary to other inferences, Modus Ponens is arguably automatic and, as long as the conditional major premise is encoded in WM, it can be triggered unconsciously by a minor premise that is below the threshold of detection. This finding is in keeping with other recent findings focused on the neural basis of human deduction that suggest that human reasoning might be a diversified, componential ability, where apparently similar logical steps can be carried out in different ways and by different neural circuits (Goel, 2007; Prado, Van Der Henst, & Noveck, 2010; Reverberi et al., 2012, 2010; Reverberi, Shallice, D'Agostini, Skrap, & Bonatti, 2009). Similarly, our findings

suggest that early stages of processing do not trigger fallacious arguments, such as the Affirmation of the Consequent, even though the same fallacious arguments might be triggered by later, and explicit, stages of reasoning. This observation is consistent with recent studies showing that even when people give illogical, belief-biased responses, for example when solving categorical syllogisms, they are nevertheless able to detect that their response is not fully warranted (De Neys & Franssens, 2009).

The present findings finally suggest a straightforward way to reconcile the contrasting pieces of evidence concerning the automaticity of the Modus Ponens. On one hand, all the studies supporting the automaticity of the Modus Ponens (including Lea, 1995, Rader & Sloutsky, 2002, and the present one) tested Modus Ponens indirectly. The participants were required neither to draw nor to evaluate a Modus Ponens argument: The dependent variables were indirect, e.g., a recognition memory task or a priming effect on a secondary tasks. On the other hand, the effects suggesting that Modus Ponens is *not* automatic (including the modulation of Modus Ponens by additional premises; e.g., Byrne, 1989; Stevenson & Over, 1995), and its varying rates in accordance to cognitive ability (e.g., Newstead, Handley, Harley, Wright, & Farrelly, 2004) concern *explicit* Modus Ponens inferences: Participants in those studies were explicitly required to evaluate whether or not a given conclusions followed from a given set of premises. This pattern of apparently contrasting pieces of evidence hints at the possibility that two mechanisms contribute to the Modus Ponens inference: a fast one automatically triggered by the logical structure alone and a slow non-automatic one that integrates other processes and that is sensitive to previous knowledge about the truth of the premises. The process ending first cues a

response, which, depending upon task instructions, available resources, and time, may need to be inhibited in favour of an alternative, less rapidly cued response (see also the "parallel competitive account" in Handley, Newstead, & Trippas, 2011). This view is also consistent with a previous interpretation of the modulation of explicit Modus Ponens by "disabling" premises proposed by Stevenson and Over (1995): additional premises and background knowledge reduce the believability of one or both of the premises of the Modus Ponens. Because the participants doubt the premises, then they also doubt the correctness of the conclusion cued by those premises, and they sometimes do not endorse it. In so doing, they confound the validity of an argument with its soundness[1]. In the present perspective, the fast, automatic, implicit mechanism that cues the acceptance of the Modus Ponens inference is sensitive to its validity alone; the "slow lane" mechanism – by integrating previous knowledge – is also sensitive to soundness, in keeping with that *belief-driven* responses are slower than some *logic-driven* responses reported by (Handley et al., 2011).

Beyond the importance of these results for further improving current theories of human deduction, they also extend our understanding of the cognitive activities that can be performed without consciousness. The existence of subliminal perception has been accepted by the scientific community. What is still a matter of controversy is both the depth of processing of subliminal stimuli and the type of cognitive processes that subliminal stimuli can trigger. Thus, although it is largely accepted that subliminal stimuli can activate lower-level

---

[1] In logic a *valid* conclusion is a conclusion that necessarily follows from a given set of premises (disregarding the factual truth of the premises). A *sound* conclusion is a valid conclusion from factually true premises.

processing (Greenwald, 1992), the possibility that subliminal stimuli can also trigger semantic (Dell'Acqua & Grainger, 1999; Naccache & Dehaene, 2001b) or inferential (e.g., deduction, as in the present study) computations is debated (for a review see Kouider & Dehaene, 2007). However, in recent years, an increasing number of studies have provided evidence of subliminal activation in high-level processing, including in decision making (Soon, Brass, Heinze, & Haynes, 2008), motivation (Pessiglione et al., 2007), emotion (Etkin et al., 2004), and in the extraction of formal features from audio streams (Peña, Bonatti, Nespor, & Mehler, 2002). The present study indicates that a critical component of deductive reasoning – the Modus Ponens – can be triggered by a non-conscious stimulus and performed without voluntary control. Thanks to this inferential tool, unconscious processing - already proven to be able to extract structural rules (e.g., Peña et al., 2002) - may also be able to use these rules by activating their predictions (i.e., the consequents) whenever the triggering conditions (i.e., the antecedents) are satisfied. In this way, unconscious processing of Modus Ponens may play the same role in unconscious cognition as deduction in explicit thinking, i.e., it may highlight the consequences that follow from previously held pieces of knowledge.

7. Conclusions

The present study provides evidence that *Modus Ponens* ("if p then q; p; therefore, q"), unlike any other valid or invalid inference that we tested (including Modus Tollens and Disjunctive Syllogism), is an *automatic* information processing step that can be performed even when individuals are

not aware of a subliminal second premise ("p"). These results suggest that this form of deductive reasoning can be included in the group of high-level cognitive tasks that are performed unconsciously. Second, these findings are not fully consistent with current cognitive theories of human deductive reasoning and, thus, may prompt further developments and revisions.

References

Braine, M. D. S., & O'Brien, D. P. (1998). *Mental logic*. Mahwah, NY, US: Lawrence Erlbaum Associates.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, *31*(1), 61–83.

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

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences*, *10*(5), 204–211.

Dell'Acqua, R., & Grainger, J. (1999). Unconscious semantic priming from pictures. *Cognition*, *73*(1), B1–B15.

Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General*, *127*(3), 286–303.

Etkin, A., Klemenhagen, K. C., Dudman, J. T., Rogan, M. T., Hen, R., Kandel, E. R., & Hirsch, J. (2004). Individual Differences in Trait Anxiety Predict the Response of the Basolateral Amygdala to Unconsciously Processed Fearful Faces. *Neuron*, *44*(6), 1043–1055.

Evans, J. S. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: the psychology of deduction*. Hove, UK: Lawrence Erlbaum Associates.

Goel, V. (2007). Anatomy of deductive reasoning. *Trends in Cognitive Sciences*, *11*(10), 435–441.

Greenwald, A. G. (1992). New Look 3: Unconscious cognition reclaimed. *American Psychologist*, *47*(6), 766–779.

Grosset, N., & Barrouillet, P. (2003). On the nature of mental models of conditional: The case of If , If then , and Only if. *Thinking & Reasoning*, *9*(4), 289–306.

Handley, S., J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 28–43.

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

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

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 857–875.

Kouider, S., & Dehaene, S. (2009). Subliminal Number Priming Within and Across the Visual and Auditory Modalities. *Experimental Psychology (Formerly 'Zeitschrift Für Experimentelle Psychologie')*, *56*(6), 418–433.

Kouider, S., & Dupoux, E. (2001). A functional disconnection between spoken and visual word recognition: evidence from unconscious priming. *Cognition*, *82*(1), B35–B49.

Lea, R. B. (1995). On-line evidence for elaborative logical inferences in text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(6), 1469–1482.

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), 227–234.

Naccache, L., & Dehaene, S. (2001a). The Priming Method: Imaging Unconscious Repetition Priming Reveals an Abstract Representation of Number in the Parietal Lobes. *Cerebral Cortex*, *11*(10), 966–974.

Naccache, L., & Dehaene, S. (2001b). Unconscious semantic priming extends to novel unseen stimuli. *Cognition*, *80*(3), 215–229.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226–254.

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

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-Driven Computations in Speech Processing. *Science*, *298*(5593), 604–607.

Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R. J., & Frith, C. D. (2007). How the Brain Translates Money into Force: A Neuroimaging Study of Subliminal Motivation. *Science*, *316*(5826), 904–906.

Prado, J., Van Der Henst, J.-B., & Noveck, I. A. (2010). Recomposing a fragmented literature: How conditional and relational arguments engage different neural systems for deductive reasoning. *NeuroImage*, *51*(3), 1213–1221.

Rader, A. W., & Sloutsky, V. M. (2002). Processing of logically valid and logically invalid conditional inferences in discourse comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 59–68.

Reverberi, C., Bonatti, L. L., Frackowiak, R. S. J., Paulesu, E., Cherubini, P., & Macaluso, E. (2012). Large scale brain activations predict reasoning profiles. *NeuroImage*, *59*(2), 1752–1764.

Reverberi, C., Cherubini, P., Frackowiak, R. S. J., Caltagirone, C., Paulesu, E., & Macaluso, E. (2010). Conditional and syllogistic deductive tasks dissociate functionally during premise integration. *Human Brain Mapping*, *31*(9), 1430–1445.

Reverberi, C., Shallice, T., D'Agostini, S., Skrap, M., & Bonatti, L. L. (2009). Cortical bases of elementary deductive reasoning: Inference, memory, and metadeduction. *Neuropsychologia*, *47*(4), 1107–1116.

Rips, L. J. (1988). Deduction. In R. J. Sternberg & E. E. Smith, *The psychology of human thought* (pp. 116–152). New York, NY, US: Cambridge University Press.

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

Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., & Pessoa, L. (2008). Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends in Cognitive Sciences*, *12*(8), 314–321.

Soon, C. S., Brass, M., Heinze, H.-J., & Haynes, J.-D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, *11*(5), 543–545.

Stevenson, R. J., & Over, D. E. (1995). Deduction from Uncertain Premises. *The Quarterly Journal of Experimental Psychology Section A*, *48*(3), 613–643.