Psychological Topics 23 (2014), 1, 143-165

Original scientific paper – UDC – 159.95.072-057.875
159.955.5.072-057.875

Metacognitive Judgments and Syllogistic Reasoning

Igor Bajšanski, Maja Močibob

Department of Psychology, Faculty of Humanities and Social Sciences,
University of Rijeka, Croatia

Pavle Valerjev

Department of Psychology, University of Zadar, Croatia

Abstract

In this study we examined the relationship between metacognitive judgments and accuracy in syllogistic reasoning. In three experiments, the participants made different types of metacognitive judgments: confidence judgments, judgments of performance and judgments of task difficulty. The confidence judgments were made (a) after producing a conclusion that logically followed from given premises (Experiment 1) and (b) after choosing the conclusion from a list of choices (Experiment 2), and judgments of performance were made (c) after a quick overview of a problem (Experiment 3). Judgments of difficulty were made before (Experiment 3) and after (Experiment 2) solving syllogistic problems. A total of 166 psychology students participated in the experiments. In all three experiments, participants were generally overconfident. The relative accuracy of judgments was generally low, with the exception of Experiment 1. The results do not indicate clearly that judgments made after solving tasks are more accurate than judgments made before solving them. The additional analysis by items showed positive correlations between different metacognitive judgments but lower correlations between judgments and reasoning accuracy, and negative correlations between metacognitive judgments and response times.

Keywords: metacognition, syllogistic reasoning, judgment accuracy

Introduction

Metacognitive processes of monitoring and control have been extensively studied in the domain of memory (Dunlosky & Bjork, 2008; Hart, 1965; Koriat, 1993, 1997, 2007; Nelson & Dunlosky, 1991; Nelson & Narens, 1990; Schwartz, 1994) and text comprehension (Baker & Brown, 1984; Glenberg & Epstein, 1985; Lin & Zabrucky, 1998; Maki, 1998; Rawson & Dunlosky, 2002; Weaver, 1990;

✉ Igor Bajšanski, Department of Psychology, Faculty of Humanities and Social Sciences, University of Rijeka, 51000 Rijeka, Sveučilišna avenija 4, Croatia. E-mail: *sibajsan@ffri.hr*

Wiley, Griffin, & Thiede, 2005). It has been recognized that it is important to expand the study of metacognition to other domains of cognition, in particular to the psychology of thinking, including problem solving (Metcalfe, 1986a, 1986b; Metcalfe & Wiebe, 1987), and reasoning, judgment and decision making (Thompson, 2009; Thompson, Prowse Turner, & Pennycook, 2011). It can be argued that the metacognitive processes of monitoring and control play important roles in the processes of thinking. However, Thompson et al. (2011) state that "surprisingly, there is little research on the role of these types of metacognitive processes in reasoning and decision making" (p. 109). In this article, we focus on the role of metacognitive monitoring in syllogistic reasoning. Although syllogistic reasoning is one of the most studied domains of human reasoning, there are not many studies of metacognitive processes in this domain.

General Framework for the Study of Metacognitive Monitoring and Control

Nelson and Narens (1990) proposed a theoretical framework for research on metacognitive monitoring and control of memory. According to this model, people make various judgments when they attempt to monitor and control their own memory such as how easy it will be to learn new material (ease of learning), how well the material is learned (judgment of learning) and judgment that we already know a certain fact (feeling of knowing). Different metamemory processes are engaged during different phases of memorization (acquisition, storage or retrieval). Modern views hold that metacognitive judgments are based on inferential processes (e. g. Koriat, 1997; Schwartz, 1994). People do not have direct access to memory traces, but their metacognitive judgments rely on various available cues. Contemporary research and theoretical work on metacognition addresses the problem as to which cues people use when they make various metacognitive judgments in different learning tasks and in which way metacognitive monitoring influences regulation of cognition.

The relationship between metamnemonic judgments and memory performance, as well between comprehension and judgments of comprehension can be determined in different ways (Benjamin & Diaz, 2008; Koriat, 2007; Schraw, 2009). Two aspects of this relationship can be distinguished: absolute accuracy and the relative accuracy of metacognitive judgments. Absolute accuracy (or calibration) is the degree to which the mean value of judgments corresponds to mean actual memory or comprehension performance. It can be evaluated with calibration curves, and measures of absolute accuracy such as the bias index. Measures of absolute accuracy indicate whether people are overconfident or underconfident. Relative accuracy (or resolution) refers to the degree to which judgments are correlated with performance across items. It indicates how well a person can discriminate learned and not learned items. Relative accuracy is measured by the within-subject correlation between judgments and performance, and the most common measure used in metacognitive experiments is Goodman-Kruskal gamma coefficient (Nelson, 1984).

The Role of Metacognition in Thinking: Metacognitive Reasoning Theory

Thompson (2009; Thompson et al., 2011) attempted to bridge the gap between the dominant approach in the study of reasoning, judgment and decision making (dual processes approach) and the study of metacognition. According to Metacognitive reasoning theory, a metacognitive judgment, called The Feeling of Rightness (FOR), accompanies outputs of the fast, automatic, and intuitive processes of reasoning (Type 1 processes). This judgment signals whether the outputs of these processes are sufficient or whether we need to engage slow and analytic processes (Type 2) and determines both the quality and the degree of engagement of Type 2 analytical processes. When a Feeling of Rightness that accompanies an initial answer is weak, the initial answer is likely to be reconsidered and changed. This relationship between FOR and analytic thinking was demonstrated by Thompson et al. (2011) in several experiments, in different domains of reasoning, judgment and decision making. In given tasks, participants first provided a fast, intuitive answer and rated their Feeling of Rightness of that answer. After that, they were allowed as much time as needed to reconsider their first answer and to provide their final answer. The weaker FOR was associated with longer rethinking time and the greater probability of changing the initial answer.

The study of metacognitive monitoring in reasoning (and also in judgment and decision making) can also be focused on more general problems, such as whether confidence is correlated with performance, how accurate the reasoners' metacognitive judgments are, can reasoners discriminate tasks which they solved correctly from those that they did not solve correctly, and is the accuracy of their metacognitive judgments related to reasoning ability. Similar problems are routinely addressed in the study of metamemory, but most of them have not been addressed so far in most domains of reasoning. In this study we examined the accuracy of metacognitive judgments in syllogistic reasoning. Before providing a description of studies on metacognitive processes in syllogistic reasoning, we must first provide a short background on syllogistic reasoning.

Syllogistic Reasoning

The processes of syllogistic reasoning have been studied in a large number of psychological studies (Bara, Bucciarelli, & Johnson-Laird, 1995; Bucciarelli & Johnson-Laird, 1999; Khemlani & Johnson-Laird, 2012). A syllogism is a form of logical argument which is comprised of three propositions: two premises and a conclusion which is derived from the premises. Every premise has its own term (category) which also forms a part of the conclusion and both premises also contain a middle term which does not form a part of the conclusion.

Categorical syllogism contains categorical propositions, which either affirm or negate that members of one category belong to another category. For example, in the proposition *All birds are animals*, it is stated that all individuals which belong to the category of birds belong to the category of animals as well. Certain propositions affirm that the members of one category belong to another category (*affirmative* propositions), while other propositions negate this (*negative* propositions). Furthermore, certain propositions place all the members of one category in a relation to another category (*universal* propositions), while other propositions place only several (*particular* propositions). Therefore, categorical propositions can appear in four types or *moods*: affirmative-universal (*All A are B*), affirmative-particular (*Some A are B*), negative-universal (*No A are B*) or negative-particular (*Some A are not B*), which are labeled as A, I, E, O (abbreviations derived from Latin words AffIrmo and nEgO).

A syllogism is valid when the conclusion necessarily follows from the premises. Every syllogism proposes a hypothetical situation which we must assume contains truthful premises, despite what we might think about them. What is of the most importance is that the conclusion follows from the premises. In instances when a conclusion derived from premises does not necessarily follow, the syllogism is invalid. From certain pairs of premises it is sometimes impossible to derive a conclusion and in cases such as these we say the conclusion does not follow. Here is an example of a valid syllogism:

All writers are gardeners.

<u>All chemists are writers.</u>

Conclusion: All chemists are gardeners.

This syllogism is valid because the derived conclusion necessarily follows from the proposed premises. An example of an invalid syllogism can be seen in the following:

No gardener is a writer.

<u>No chemist is a writer.</u>

Conclusion: No chemist is a gardener.

This syllogism is invalid because the derived conclusion does not necessarily follow from the proposed premises. Furthermore, this syllogism does not allow one to derive a valid conclusion.

Syllogisms may appear in four different figures. These figures show the various distributions of the middle term which occurs in both premises (B) and the end terms which occur in the conclusion (A and C). The four figures are:

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

In the previous example, the terms in the premises are in the following figure:

A-B (gardeners-writers)
C-B (chemists-writers)

Every premise may take the form of any of the four moods of categorical proposition (A, I, E, O). The number of possible pairs of premises for every figure amounts to 16, which means that for four figures we have a total of 64 pairs of premises. Furthermore, there are eight possible conclusions to the 64 pairs of premises: the four A-C moods and four C-A moods. This adds up to a total of 512 syllogisms. However, out of 64 pairs of premises, 27 yield at least one valid conclusion, whereas the remaining 37 syllogistic forms are invalid (Bucciarelli & Johnson-Laird, 1999; Khemlani & Johnson-Laird, 2012).

Numerous psychological studies identified a variety of factors that influence the accuracy of syllogistic reasoning (for example, atmosphere and believability). Psychological experiments on syllogistic reasoning have used different methods (Geurts, 2003). Participants may be asked to choose from a list of possible conclusions (multiple-choice paradigm), to write a conclusion (production paradigm), or to evaluate a given syllogism (evaluation paradigm). These three paradigms yield very similar results. One of the most stable finding of these studies is that syllogisms differ in difficulty in a systematic way. Some syllogisms are very easy, and some syllogisms are very hard for untrained adults. For example, Khemlani and Johnson-Laird (2012) analyzed the results of six studies and reported that the percentage of correct conclusions for 64 syllogisms goes from 1% for the most difficult syllogism (*All B are A, None of the C is B*) to 90% for the easiest syllogism (*All the B are A, Some of the C are B*). Typically, in easy syllogisms reasoners give one or two predominant responses, while in difficult syllogisms reasoners have a greater variety of responses.

Studies of Metacognition in Syllogistic Reasoning

Quayle and Ball (2000, Experiment 1) tested the effect of validity of syllogisms on confidence ratings. They obtained higher acceptance rate for valid than for invalid syllogisms, but the acceptance rate for invalid syllogisms was nonetheless high (73%). Their participants were generally confident in their answers, but had higher confidence ratings for valid than for invalid syllogisms.

Thompson and her colleagues conducted several experiments aimed to examine the role of metacognition in syllogistic reasoning (Prowse Turner & Thompson, 2009; Shynkaruk & Thompson, 2006). They examined the participants' confidence judgments of their performance in syllogistic reasoning tasks and noticed the existence of overconfidence displayed in the dissociation between actual accuracy of performance and confidence in that performance.

Shynkaruk and Thompson (2006) conducted two experiments using syllogistic reasoning problems that varied in the believability and validity of their conclusions and showed very small or nonexistent correlations between the reasoners' confidence and accuracy. The participants were presented with 12 syllogisms consisting of two premises which were followed by a conclusion. Their task was to judge the validity of the presented conclusion, by responding *yes* or *no*, and to estimate their confidence in the given answer on a 7-point confidence scale. Of the two premises used in each task, one was of the type: *No _ are _* and the other was of the type *Some _ are _*. The conclusion which followed the premises was in the form of *Some _ are not _*. The participants answered each problem twice; first within a deadline of 10 seconds and then after one minute in which they had time to reconsider their initial answer. A and C terms in syllogisms referred to familiar categories and the B term to nonsense categories for believable problems, while A and C terms were nonsense terms and the B term referred to familiar categories for neutral problems. There were two conclusion forms: *Some C are not A* and *Some A are not C*. The results of both experiments indicated confidence in the accuracy of a given answer as a poor indicator of the actual accuracy of the performance. Two factors that impact confidence were identified: the believability of the problem and the time available for responding. Confidence was greater for believable and unbelievable conclusions rather than neutral and it was lower for problems that had to be solved within a shorter time limit. The accuracy was greater only for unbelievable conclusions. Although both confidence and accuracy increased when participants were allowed more time for answering, the increase in confidence was independent of the increase in accuracy. Confidence was not well calibrated with performance, but participants were more likely to change their answers when confidence in their initial answer was low, regardless of whether that answer was correct or not.

According to Prowse Turner and Thompson (2009), confidence judgments for conclusions in syllogistic reasoning can be based on three factors. One factor is the Feeling of Rightness, which refers to the metacognitive cue that the answer is correct and that there is no need to look further for the answer (Thompson, 2009). Another factor on which confidence judgments in syllogistic reasoning can be based, refers to the external aspects of the environment, such as time limitations. The third factor they identified as important for confidence judgments are individual differences. They conducted the experiment in which participants had to assess the validity of inferences of syllogistic problems (to indicate whether the conclusion that follows from the given premises was necessary, possible or impossible) and to provide estimates of their confidence in the given answer. They used training as a between-participants variable (half of the participants received training in solving syllogistic problems prior to completing the test problems and the other half did not) and the number of mental models required to represent the problem and conclusion type as within-participant variables. Training improved the participants' accuracy and it increased the participants' ability to estimate the number of correct answers although it had no effect on their overall levels of confidence. The number of models required to represent the problem and logical necessity showed systematically different relationships to confidence and accuracy. Although there was no difference in accuracy, the participants' confidence was greater in single-model problems than in multiple-model problems.

Aims of the Study and Overview of Experiments

Described studies focused on the relation between confidence and accuracy, and we wanted to expand on these studies in two ways.

First, multiple metacognitive judgments were used in the experiments. General questions about the relationship between confidence and accuracy can be extended to other types of metacognitive judgments as well. In the study of metamemory, confidence judgments are one type of the metacognitive judgments given after producing an answer. Other types of metamnemonic judgments have also been extensively studied, such as judgments of learning. These types of judgments are made before producing an answer in a memory test, after studying. In a similar vein, we focused on metacognitive monitoring processes that occur before the actual solving of syllogistic problems. Furthermore, another type of metacognitive judgments was also examined – judgments of difficulty. Thus, we used four different judgments: confidence after solving syllogistic tasks, judgments of performance made before solving syllogistic tasks, and judgments of difficulty made before and after solving, and we examined their relations to reasoning accuracy.

Second, syllogistic problems with different levels of difficulty from all the four figures were used as stimuli in order to avoid the problem of restricted range. Calculated judgment accuracy depends on the range of performance scores (Schwartz & Metcalfe, 1994): If all items about which participants make metacognitive judgments are on a similar level of difficulty (restricted range), the correlation between judgments and accuracy will be lower than with more various items. Shynkaruk and Thompson (2006), Prowse Turner and Thompson (2009) and Thompson et al. (2011) used homogenous sets of syllogisms in their experiments. For example, Shynkaruk and Thompson (2006) used only syllogisms which consisted of one universal negative premise and one particular affirmative premise, and a particular negative conclusion. The greater variability of syllogisms with different levels of difficulty could result in higher correlations between judgments and performance.

The main aim of the present study was to examine metacognitive judgments, their accuracy and their relation to the performance in syllogistic reasoning tasks. In addition, we focused on several specific aims. The first aim was to analyze the correlation between metacognitive judgments and reasoning accuracy. On the basis of previous studies, the hypothesis was that judgments and accuracy are not correlated, or that the correlation is low. Furthermore, we expected the reasoners to be overconfident, as it was shown in many domains of cognition. The second aim was to analyze the relative accuracy of metacognitive judgments measured by the Goodman-Kruskal gamma coefficient, which is a standard procedure in the study of metamemory and metacomprehension. This is a measure which indicates whether reasoners can discriminate between tasks which they solved and did not solve correctly. We expected that participants would show low relative accuracy, because low relative accuracy is typically found in complex cognitive domains such as reading comprehension. The third aim was to examine the relationship between relative judgment accuracy and reasoning accuracy. We expected this correlation to be low.

Three experiments were conducted. In all three experiments, we applied the basic procedure used by Shynkaruk and Thompson (2006) and Quayle and Ball (2000), that is, we collected the metacognitive judgments on a 7-point Likert type scale. In Experiment 1 and Experiment 2 we studied metacognitive judgments made after solving syllogistic problems (confidence and judgments of difficulty). In Experiment 1 the tasks were presented in a paper-and-pencil form, and in Experiment 2 and 3 the stimuli were presented on a computer screen, and measures of the time of processing were also used. In Experiment 3, we examined metacognitive judgments made before solving syllogistic problems. Participants were treated in accordance with the APA Ethical Guidelines.

Experiment 1

The aim of this paper-and-pencil pilot study was to test confidence judgments on a representative sample of valid and invalid syllogisms from all figures and with different levels of difficulty. The participants were provided with 24 syllogistic reasoning problems that varied with respect to validity, figures and difficulty. For each problem, they were asked to write down the conclusion that follows logically from two premises as well as to provide a confidence judgment in the given answer.

Method

Participants

A total of 40 psychology students (2 males, 38 females) from the University of Rijeka took part in this study in exchange for course credit.

Materials

Twenty-four syllogistic reasoning problems were presented in a booklet form. Each problem consisted of two premises. A, B and C terms in premises referred to different professions; each pair of premises contained a unique combination of professions. The participants were instructed to write down a conclusion that logically follows from the two given premises or to write *no valid conclusion* if they thought it was impossible to deduct a logically valid conclusion from the given premises. This was followed by a 7-point scale in which they were to express their confidence in the correctness of their answer, where 1 corresponded to *not at all confident* and 7 corresponded to *extremely confident*.

The syllogistic problems varied with respect to validity, figures and difficulty. We included sixteen syllogisms with a logically valid conclusion, and eight invalid syllogisms. The proportion of possible figures was equal across syllogisms (4 valid and 2 invalid syllogisms in each figure). Levels of difficulty across syllogisms ranged from 16% to 88% of correct responses, according to Khemlani & Johnson-Laird (2012).

Procedure

Participants were tested in groups. No time limit was set for completing the problems. Testing took approximately 20 minutes.

Each participant received a booklet which consisted of 24 syllogism problems, 4 problems per page. Half of the participants received the booklet A and the other half received the booklet B, containing the same problems presented in different order. Following instructions were presented on the first page of the booklets:

> This experiment examines how people solve logical problems. Your task is to write a conclusion that you think follows logically from the two given premises. For example, two premises are given:
>
> All people are mortal.
> All artists are people.
> Conclusion: ________________________

Which conclusion follows logically from the two premises? The correct answer is *All artists are mortal.* Write down your answer. Below the space for the answer, there is a seven-point scale. Your task is to rate your confidence in the correctness of your answer: 1 means *not confident at all* and 7 means *very confident*.

Please, keep in mind two important things while solving the tasks. First, while answering, you must assume that all that is stated in two premises is true. Namely, if a task contains a premise such as *All merchants are writers,* you must assume that this premise is true, regardless of whether the premise is in fact true or not and regardless of your own opinion on whether this premise is true or not.

Secondly, it is important to keep in mind that it is not always possible to deduct a conclusion from given premises. For example, two premises are given:

No flowers are animals.

No rocks are animals.

No conclusion follows logically from the two premises and in this case you can write down *no valid conclusion* in the space for the answer and rate your confidence in the answer on a seven-point scale.

Finally, please, complete all tasks.

Results

Reasoning Accuracy

The answer was scored as correct when it was one of the valid conclusions for the given syllogism, or when the answer was *no valid conclusion* for syllogisms for which there is no valid conclusion (Khemlani & Johnson-Laird, 2012).

The mean number of correct answers was 12.23 (SD=4.75). The mean percentage of correct answers for valid syllogisms was 52%, and for invalid 49%. Thus, the proportion of correct answers was not statistically different due to the validity of syllogisms (*Wilcoxon* Z=.73, p>.05).

Confidence

The average judgments and relative judgment accuracy obtained across the three experiments are presented in Table 1. Confidence ratings were averaged over correct and incorrect responses. The mean confidence for correct responses was 5.15 (SD=1.01), and for incorrect 4.61 (SD=1.01). The difference was significant [t(39)=5.86, p<.01]. However, it is important to note that the average confidence for incorrect responses was relatively high-above the middle point on a seven-point rating scale. Table 2 presents the proportions of correct answers for each value of self-reported confidence across all participants in three experiments. Although the bias index or some other index of absolute accuracy cannot be computed with this type of rating scale, the obtained results clearly point to overconfidence. Out of 469 total incorrect responses, 31% were given high confidence ratings (values 6 and 7), and 23% were given rating 5.

Table 1. *Average Metacognitive Judgments and Relative Judgment Accuracy in Three Experiments*

Experiment and type of judgment		Ratings overall	Ratings of correct responses	Ratings of incorrect responses	Gamma
Experiment 1:					
Confidence	*M*	4.94	5.15	4.61	.31
	SD	1.00	1.01	1.01	.35
Experiment 2:					
Confidence	*M*	4.99	5.05	4.91	.13
	SD	1.30	1.39	1.30	.52
Difficulty	*M*	3.10	3.05	3.18	-.03
	SD	1.10	1.17	1.10	.45
Experiment 3:					
Performance	*M*	5.10	5.08	5.07	.07
	SD	1.02	1.13	0.98	.36
Difficulty	*M*	3.21	3.19	3.27	-.10
	SD	0.92	1.03	0.89	.36

Table 2. *Total Number of Responses (N) and Proportions of Correct Responses for Each Value of Metacognitive Judgments in Three Experiments*

Experiment and type of judgment		Ratings						
		1	2	3	4	5	6	7
Experiment 1:								
Confidence	*N*	18	56	109	171	213	206	183
	Accuracy	.33	.27	.40	.43	.49	.62	.63
Experiment 2:								
Confidence	*N*	26	39	96	128	139	146	194
	Accuracy	.08	.46	.34	.31	.42	.35	.50
Difficulty	*N*	114	197	205	95	91	43	23
	Accuracy	.46	.39	.43	.39	.23	.35	.39
Experiment 3:								
Performance	*N*	20	23	62	148	159	124	184
	Accuracy	.40	.52	.44	.42	.40	.51	.48
Difficulty	*N*	120	160	149	150	106	43	16
	Accuracy	.44	.54	.41	.44	.42	.51	.44

For each participant the Goodman-Kruskal gamma correlation coefficient between confidence ratings and response accuracy was calculated. The average gamma was .31, and it was significantly different from zero, as it was revealed by one-sample t-test [$t(39)=5.66$, $p<.01$].

Confidence and Reasoning Accuracy

In order to analyze whether metacognitive judgments are related to reasoning accuracy, two Pearson correlation coefficients were computed. First, the correlation between reasoning accuracy (measured as the number of correctly solved reasoning problems) and the confidence ratings averaged across all items for each participant was .25, $p=.06$, one-tailed. Second, the correlation between the relative accuracy of metacognitive judgments (gamma coefficient computed for each participant) and reasoning accuracy was .52, $p<.001$, one-tailed. Thus, good reasoners showed a tendency to assign higher ratings to their performance (although this effect was not significant) than poor reasoners, but they also were better at discriminating between items which they solved correctly from those which they solved incorrectly.

Experiment 2

Experiment 2 was a conceptual replication of Experiment 1: on a representative sample of syllogistic problems, participants were asked to make metacognitive judgments after solving each problem. The experiment differs from Experiment 1 in several ways. First, the tasks were presented on a computer screen. Second, a different sample of syllogisms was used, created in a similar way as in Experiment 1. Third, instead of open-ended problems, multiple choice problems were used: participants had to choose a correct answer among five alternatives. With this experiment we wanted to test whether the results obtained in Experiment 1 can be replicated and generalized.

Method

Participants

Sixty four psychology students (15 males, 49 females) from the University of Zadar took part in this experiment in exchange for course credit. They were randomly and equally assigned to one of two groups: confidence and judgment of difficulty group.

Materials

Syllogisms. Twenty four syllogism problems were used in this study. In each trial, participants were shown two premises containing professions as A, B and C terms. There was a unique combination of three professions in each task. Sixteen of the used problems were valid and eight were invalid. All four possible figures for the premises were used equally across tasks, so each figure was used in six tasks (in four valid and two invalid syllogisms). Task difficulty ranged from 3% to 82% of correct answers (Khemlani & Johnson-Laird, 2012). Each syllogism was followed by five possible answers, that is, four conclusion statements and a *no valid conclusion* option. The conclusion statements were presented in two forms. For half of the valid and half of the invalid syllogisms the given conclusions were in AC form: 1) *All A are C*, 2) *Some A are C*, 3) *No A are C*, 4) *Some A are not C*, 5) *No valid conclusion*; for the other half the given conclusions were in CA form: 1) *All C are A*, 2) *Some C are A*, 3) *No C are A*, 4) *Some C are not A*, 5) *No valid conclusion*. Participants selected their response by pressing the correspondent key on the keyboard. Six of the twenty four syllogisms chosen for this experiment were syllogisms with multiple possible logically valid conclusions. For those, we presented the four given conclusions in a form in which there was only one valid conclusion. For example, a syllogism which consists of the two following premises: *All miners are architects and All architects are biologists*, has the following valid conclusions (Khemlani & Johnson-Laird, 2012): *All miners are biologists*, *Some miners are biologists*, *Some biologists are miners*. In this task we presented the given answers in CA form to assure that there is only one (*Some biologists are miners*) correct conclusion among the given conclusions.

Metacognitive Judgments. After solving each reasoning problem, the participants were asked to make a rating on a 7-point scale. One half of the participants rated how confident they were that they solved the task correctly, where 1 means *not confident at all* and 7 means *extremely confident*, and the other half rated the perceived difficulty of the task, where 1 means *not difficult at all* and 7 means *very difficult*.

Procedure

The participants were tested individually. This experiment was programmed and run using E-prime (Schneider, Eschman, & Zuccolotto, 2002). At the beginning, instructions were presented on a computer screen. As in Experiment 1, participants were introduced to the task by explaining the logic of syllogisms using the examples of one valid and one invalid syllogism. The procedure was explained to the participants as well as how to give their answers and judgments for each task. The instructions were shown on the screen in several parts and the participants continued with the next part of the instructions by pressing the space bar. They were given as much time as they needed to comprehend the instructions.

After the instructions, the participants were given two practice problems. All of the participants received the same practice problems, one with a valid conclusion (*some C are A*) and one invalid syllogism with no logically valid conclusion.

After the practice problems the participants began solving the remaining 24 syllogisms. They pressed the space bar to display each task. First, two premises with five answers, marked with numbers from 1 to 5, were shown on the screen. The participants had to choose the one that follows logically from the given premises or to choose the *No valid conclusion*, by pressing the corresponding key on the keyboard. Timing began when the problem appeared on the screen and ended when the answer (1 to 5) was given. There was no time limit for giving the answer. Second, participants made metacognitive judgments. A 7-point scale appeared on the screen. Half of the participants made confidence judgments, and half made judgments of difficulty, by pressing the corresponding key. There was no time limit for giving their estimation.

Each participant was presented with 24 randomly ordered tasks. Testing lasted an average of 20 minutes.

Results

Reasoning Accuracy

The mean number of correct answers was 9.38 (SD=3.29). The mean percentage of correct answers for valid syllogisms was 41%, and for invalid 35%. There was no difference in the proportion of correct answers for valid and invalid syllogisms (*Wilcoxon* Z=1.73, p>.05).

Metacognitive Judgments

The values of metacognitive judgments were averaged over correct and incorrect responses. The mean judgment of confidence for correct responses was 5.05 (SD=1.39), and for the incorrect 4.91 (SD=1.30). The mean judgment of difficulty for correct responses was 3.05 (SD=1.17), and for the incorrect 3.18 (SD=1.10). Neither type of judgments differed significantly between correct and incorrect responses. Considering judgments of confidence, 41% of incorrect responses were given high confidence ratings (values 6 and 7). For judgments of difficulty, 38% of wrong answers were given low judgments of difficulty (values 1 and 2). Therefore, we can conclude that participants were overconfident in their judgments (Table 2).

For each participant the Goodman-Kruskal gamma correlation coefficient between metacognitive judgments and response accuracy was calculated. The average gamma was .13 (SD=.52) for judgments of confidence, and it was -.03 (SD=.45) for judgments of difficulty. It is important to note that for judgments of difficulty negative values of gamma indicate a positive relation between the accuracy and the perceived easiness of the task. Relative accuracy of neither type of judgments, measured by the gamma index, differed significantly from zero. Therefore, our participants showed an overall inability to discriminate between items which they solved correctly from those which they solved incorrectly.

Metacognitive Judgments and Reasoning Accuracy

The Pearson correlation coefficients between reasoning accuracy and average judgments and the corresponding gamma coefficients were computed for judgments of confidence and judgments of difficulty. For judgments of confidence, the correlation between reasoning accuracy and average judgments was .34, $p<.05$, and the correlation between reasoning accuracy and relative judgment accuracy (gamma coefficients) was .29, $p=.05$, one-tailed tests. For judgments of difficulty the correlation between reasoning accuracy and judgments was -.23, $p>.05$, and the correlation between reasoning accuracy and relative judgments accuracy (gamma coefficients) was -.40, $p<.05$, one-tailed tests. Although relative accuracy was low overall, it seems to be related to reasoning accuracy at least to some degree.

Experiment 3

Experiment 3 used the same basic procedure as Experiment 2, with one exception. Instead of collecting metacognitive judgments after solving syllogistic problems, the participants were asked to make these judgments before solving the tasks and after brief exposure to the syllogistic premises. Two premises were presented on the screen for 10 seconds, and after that the participants were asked to make metacognitive judgments. In the final phase of each trial, the premises were shown again, and the task was to choose the correct answer among five options. The aim was to examine whether initial judgments are related to the performance on reasoning problems. We used a 10 seconds presentation because this interval was used in Shynkaruk and Thompson's study (2006). They explained this 10 seconds deadline as enough time to read and comprehend the premises and conclusions, yet not long enough for further deliberation.

Method

Participants

Sixty two psychology students (7 males, 55 females) from the University of Rijeka, took part in this experiment in exchange for course credit. They were randomly and equally assigned one of two conditions: judgment of performance and judgment of difficulty group.

Materials

The same 24 syllogistic reasoning problems used in Experiment 2 were used in Experiment 3.

Procedure

The procedure was similar to the procedure used in Experiment 2. However, instead of making metacognitive judgments after solving reasoning problems, the participants were asked to make metacognitive judgments before solving problems. First, for each task, two premises appeared on a computer screen for 10 seconds.

After that, a 7-point scale appeared on the screen. Half of the participants were asked to rate their confidence that they will give the correct answer to that task by pressing the corresponding key and the other half of the participants were asked to rate the difficulty of the task by pressing the corresponding key. There was no time limit for giving the judgments.

Second, two premises appeared on a screen again, but this time with five possible answers and, by pressing the corresponding key on the keyboard, the participants had to choose a conclusion that follows logically from the given premises or to choose *No valid conclusion*. Timing began when a problem appeared on the screen and ended when the answer (1 to 5) was given. There was no time limit for giving the answer.

Instructions from Experiment 2 were modified and adapted to correspond to this procedure. Each participant was presented with 24 randomly ordered tasks. Testing lasted an average of 20 minutes.

Results

Data from one participant were excluded from all analyses due to extremely low reaction times on three tasks.

Reasoning Accuracy

The mean number of correct answers was 10.90 (SD=3.56). The mean percentage of correct answers for valid syllogisms was 48%, and for the invalid 40%. The proportion of correct answers was higher for valid than for invalid syllogisms (*Wilcoxon* Z=2.18, p<.05).

Metacognitive Judgments

The values of metacognitive judgments were averaged over correct and incorrect responses. The mean judgment of performance for correct responses was 5.08 (SD=1.13), and for the incorrect 5.07 (SD=0.98). The mean judgment of difficulty was 3.19 (SD=1.03) for correct responses, and 3.27 (SD=0.89) for the incorrect. Neither type of judgments differed significantly between correct and incorrect responses. Similarly to Experiments 1 and 2, participants were overconfident in both judgments of performance and judgments of difficulty (see Table 2).

For each participant the Goodman-Kruskal gamma correlation coefficient between metacognitive judgments and response accuracy was calculated. The average gamma was .07 (*SD*=.36) for judgments of performance, and it was -.10 (*SD*=.36) for judgments of difficulty. It is important to note that for judgments of difficulty, negative values of gamma indicate a positive relation between the accuracy and the perceived easiness of task. The relative accuracy of neither type of judgments, measured by the gamma index, differed significantly from zero.

Metacognitive Judgments and Reasoning Accuracy

For judgments of performance, neither average judgments (r=.18, p>.05), neither relative accuracy (r=.29, p>.05), were significantly related to reasoning accuracy (one-tailed tests). For judgments of difficulty, the correlations between reasoning accuracy and both average judgments (-.02) and gamma coefficients (-.07) were insignificant.

We can conclude that a) the overall relative accuracy of both types of metacognitive judgments was not different from zero, and b) the relative judgment accuracy was not related to reasoning accuracy. Before the solving and after the initial inspection of the task, participants were unable to discriminate tasks for which they produced correct responses from those for which they produced incorrect responses.

Correlations Between Judgments, Reasoning Accuracy, and Response Times

Experiments 2 and 3 included the same syllogisms. For each syllogism, we computed the average data across participants: average reasoning accuracy, average response time, and average metacognitive judgments. On data obtained in this way, we performed item-based correlation analysis (Table 3). Some important results were revealed. First, reasoning accuracy was generally not related to metacognitive judgments. One exception is the significant negative correlation between reasoning accuracy in Experiment 2 and judgments of difficulty in Experiment 2 (r=-.41, p<.05): more difficult items were judged as more difficult than easier items. Judgments of performance, confidence judgments and judgments of difficulty made before solving were not related to reasoning accuracy in the analysis by items. Second, all metacognitive judgments were related, and all correlation coefficients were higher than .69. In both experiments judgments of performance and confidence were highly negatively correlated with judgments of difficulty. Therefore, syllogisms that were perceived as easy also received high confidence and performance ratings. Third, metacognitive judgments were significant predictors of response times: response times in both experiments were significantly correlated with all metacognitive judgments (the lowest r coefficient was .62). The correlations were positive for judgments of difficulty and negative for confidence judgments and judgments of performance. Thus, items that were perceived as difficult, and which induced low confidence ratings, took more time to solve.

Response times were also related to reasoning accuracy (correlation coefficients were in range from .30 to .45), but these correlations were lower than correlations with metacognitive judgments.

Table 3. *Item-Based Correlations Between Reasoning Accuracy, Response Times, and Metacognitive Judgments*

	Experiment 2				*Experiment 3*			
	Accuracy	RT	Confidence	Difficulty	Accuracy	RT	Performance	Difficulty
Experiment 2:								
Accuracy	-							
RT[a]	-.38	-						
Confidence	.14	-.73**	-					
Difficulty	-.41*	.82**	-.87**	-				
Experiment 3:								
Accuracy	.94**	-.30	.01	-.27	-			
RT[a]	-.45*	.88**	-.73**	.73**	-.40	-		
Performance	.18	-.62**	.86**	-.83**	.02	-.65**	-	
Difficulty	-.16	.65**	-.80**	.69**	-.04	.74**	-.89**	-
M	0.39	26.74[a]	4.99	3.09	0.45	17.30[a]	5.10	3.21
SD	0.25	4.75	0.32	0.47	0.27	4.16	0.61	0.58

Note. [a] Response times measured in seconds
*p<.05, **p<.01.

General Discussion

In this study, we conducted three experiments with the aim of examining metacognitive judgments made before and after solving syllogistic reasoning tasks, i.e. producing or choosing the conclusion that logically follows from given premises. Specifically, we focused on two types of judgments: judgments which are formed after solving particular syllogisms and judgments which are formed before the actual solving, after a quick overview of tasks. Furthermore, two types of judgments were used in both phases of task solving: judgments of performance/confidence and judgments of task difficulty. Furthermore, we used a wide range of task difficulty to avoid the problem of restricted range (Schwartz & Metcalfe, 1994). Relative accuracy of judgments can depend on various methodological factors, such as the number of tasks, their difficulty and range. These factors should be considered while making generalizations about the accuracy of metacognitive judgments in a certain domain. Several main results were obtained.

First, both absolute and relative accuracy of metacognitive judgments were low, and similar results were obtained by Shynkaruk and Thompson (2006), who reported that the correlations between confidence and accuracy were small or nonexistent. Regarding absolute accuracy, the results obtained in all three experiments clearly indicate that our participants were overconfident. A significant proportion of problems which were not solved correctly were assigned high metacognitive ratings, and this was observed both for judgments made before and after solving. Thus, this result fits into a larger body of research demonstrating overconfidence effects in different domains of cognition (e. g. Baranski & Petrusic, 1999; Lichtenstein, Fischhoff, & Phillips, 1982; Stankov & Crawford, 1997).

Relative accuracy of judgments was generally low. The exception from this general pattern was Experiment 1, in which participants judged their confidence after solving syllogisms and in which they solved syllogistic problems in an open-ended format. The average gamma was .31. Therefore, our participants showed a certain degree of relative accuracy, that is, the ability to differentiate between tasks which were solved correctly and those that were not solved correctly. The results of Experiment 2 and Experiment 3 also support the idea that the type of tasks used could be an important factor. In these experiments participants had to choose the correct answer from a list of choices and the average measures of relative accuracy were not significantly different from zero. Similarly, higher level of relative judgment accuracy was obtained for recall than for multiple-choice tasks in the metamemory research (Schwartz & Metcalfe, 1994). It is possible that the open-ended format improved metacognitive judgment accuracy, because the reasoning process generated more valid cues for metacognitive judgments. However, this hypothesis requires further research.

Second, the obtained results do not indicate clearly that judgments made after solving tasks are more accurate than judgments made before solving. Although this finding is common in metacomprehension literature (Maki & Serra, 1992; Pierce & Smith, 2001), the present study does not provide a conclusive effect for the "postdiction superiority effect". However, judgments made before solving the tasks, and after a quick overview of the tasks, are incorrect (Experiment 3) while judgments made after solving the tasks showed a moderate level of relative accuracy (Experiment 1). Even though relative judgment accuracy in Experiment 2 was equally low as in Experiment 3, in the analysis of interindividual correlations metacognitive judgments showed low, but positive correlations with performance, indicating a certain degree of accuracy of metacognitive monitoring.

Third, relative judgment accuracy was correlated to reasoning accuracy. Therefore, good reasoners were more able to discriminate correct from incorrect responses than poor reasoners, but only after actually solving the problems (Experiments 1 and 2).

Fourth, analysis by items showed significant correlations between different judgments, while correlations between judgments and accuracy were lower. Accordingly, syllogisms differ in their associated judgments of difficulty, which are not correlated with actual solving accuracy. Also, judgments of performance and easiness are negatively correlated with the response times: the tasks judged as solvable and easy, were solved faster, although they were not necessarily solved correctly. On the other hand, correlation between accuracy and solving time was lower. These results support Thompson et al. (2011) metacognitive reasoning theory. According to this theory, initial metacognitive judgment produces a Feeling of Rightness which mediates the quality and the extent of analytic thinking (Type 2 processing). Low FORs are associated with longer rethinking time and the probability of changing answers. However, the time reasoners spend reanalyzing their first answer will not necessarily increase the probability of giving a normatively correct answer.

The obtained correlations between different metacognitive judgments and between judgments and reasoning accuracy can also be related to the choice-independent-confidence effect (Koriat, 2008). Koriat calculated the mean confidence for correct and incorrect answers for each of 104 general knowledge questions, and found that mean confidence for correct answers correlated .64 with mean confidence for incorrect answers. Therefore, the confidence rating that is given to an item has a tendency to be high or low, independent of the correctness of the answer. Koriat interpreted this effect as a consequence of making confidence judgments on the basis of domain familiarity and by the tendency of a question to bring to mind either few or many thoughts. In our study, syllogisms differed reliably in their average metacognitive judgments, but these judgments were generally not related to accuracy. Certain syllogisms tend to induce high or low metacognitive judgments, and factors that influence judgments differ from factors that affect accuracy. Further studies should investigate these factors.

Finally, the present study has several limitations. First, although we used a representative sample of syllogisms it is important to note that not all syllogistic forms were used in the experiments. It is possible that with different sets of stimuli calculated judgment accuracy would also differ. Second, believability of syllogisms, a factor which affects both confidence and accuracy was not manipulated in this study. Third, given the differences in judgment accuracy obtained for open-ended and multiple-choice problems, further studies should be designed more carefully in order to control and compare the effects of type of tasks on metacognitive judgments.

We can conclude that metacognitive judgment accuracy in a domain of syllogistic reasoning is in part related to two methodological factors: the range of difficulty of tasks used and the format of the task itself (open-ended or multiple-choice questions). However, using a broad range of problems and open-ended problems resulted only in modest judgment accuracy. The reason for this is possibly that the problems that are not answered correctly often induce high ratings. Certain problems seem easy and solvable, independently of the actual accuracy. The basis for this overconfidence should be addressed in further research.

References

Baker, L., & Brown, A.L. (1984). Metacognitive skills and reading. In P.D. Pearson (Ed.), *Handbook of reading research* (pp. 353-394). New York: Longman.

Bara, B.G., Bucciarelli, M., & Johnson-Laird, P.N. (1995). Development of syllogistic reasoning. *The American Journal of Psychology, 108,* 157-193.

Baranski, J.V., & Petrusic, W.M. (1999). Realism of confidence in sensory discrimination. *Perception & Psychophysics, 61*, 1369-1383.

Benjamin, A.S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R.A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 73-94). New York: Psychology Press.

Bucciarelli, M., & Johnson-Laird, P.N. (1999). Strategies in syllogistic reasoning. *Cognitive Science*, *23*(3), 247-303.

Dunlosky, J., & Bjork, R.A. (2008). The integrated nature of metamemory and memory. In J. Dunlosky & R.A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 11-28). New York: Psychology Press.

Glenberg, A.M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 702-718.

Geurts, B. (2003). Reasoning with quantifiers. *Cognition*, *86*, 223-251.

Hart, J.T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*, 208-216.

Khemlani, S., & Johnson-Laird, P.N. (2012). Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, *138*(3), 427-457.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*(4), 609-639.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349-370.

Koriat, A. (2007). Metacognition and consciousness. In P.D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289-325). Cambridge: Cambridge University Press.

Koriat, A. (2008). When confidence in a choice is independent of which choice is made. *Psychonomic Bulletin & Review*, *15*(5), 997-1001.

Lichtenstein, S., Fischhoff, B., & Phillips, L.D. (1982). Calibration of probabilities: The state of the art in 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-333). Cambridge, England: Cambridge University Press.

Lin, L.-M., & Zabrucky, K.M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology, 23*(4), 345-391.

Maki, R.H. (1998). Test predictions over text material. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117-144). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Maki, R.H., & Serra, M. (1992). The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 116-126.

Metcalfe, J. (1986a). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(2), 288-294.

Metcalfe, J. (1986b). Premonitions of insight predict impending error. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(4), 623-634.

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition*, *15*(3), 238-246.

Nelson, T.O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109-133.

Nelson, T.O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "Delayed-JOL Effect." *Psychological Science*, *2*(4), 267-270.

Nelson, T.O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125-169). New York: Academic Press.

Pierce, B., & Smith, S. (2001). The postdiction superiority effect in metacomprehension of text. *Memory & Cognition*, *29*(1), 62-67.

Prowse Turner, J.A., & Thompson, V.A. (2009). The role of training, alternative models, and logical necessity in determining confidence in syllogistic reasoning. *Thinking & Reasoning*, *15*(1), 69-100.

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.

Rawson, K.A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 69-80.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburgh: Psychology Software Tools, Inc.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, *4*(1), 33-45.

Schwartz, B.L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review*, *1*(3), 357-375.

Schwartz, B.L., & Metcalfe, J. (1994). Methodological problems and pitfalls in the study of human metacognition. In J. Metcalfe & A.P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 93-113). Cambridge, MA: MIT Press.

Shynkaruk, J.M., & Thompson, V.A. (2006). Confidence and accuracy in deductive reasoning. *Memory & Cognition*, *34*(3), 619-632.

Stankov, L., & Crawford, J.D. (1997). Self-confidence and performance on tests of cognitive abilities. *Intelligence, 25,* 93-109.

Thompson, V.A. (2009). Dual-process theories: A metacognitive perspective. In J.S.B.T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171-195). New York, NY, US: Oxford University Press.

Thompson, V.A., Prowse Turner, J.A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63*(3), 107-140.

Weaver, C.A. (1990). Constraining factors in calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 214-222.

Wiley, J., Griffin, T.D., & Thiede, K.W. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology*, *132*(4), 408-428.

Received: January 12, 2014

PRAZNA PARNA STRANICA