Pavle Valerjev, Marin Dujmović

METACOGNITIVE JUDGMENTS DURING SOLVING OF WASON SELECTION TASK

Abstract

The Wason selection task is a classic problem in the psychology of reasoning. The majority of Wason selection task solvers give the wrong answer. This effect is explained by a particular cognitive bias which is called confirmation bias. Experiments usually show that this bias is strongest when the content of the task is abstract, and that when concrete content is included in the task the effect of the bias becomes lesser. According to the dual process theory, people use two different systems of thinking: Type 1 which is swift, intuitive and based on heuristics, and Type 2 which is slow, analytical and based on mental skills. The activation of the confirmation bias represents the Type 1 system of thinking. It was expected that this kind of answer would be fast, biased and that at the same time participants would have high metacognitive judgments of confidence in validity of their answers. The aim of the research was to examine efficiency and metacognitive judgments of confidence in Wason tasks. The experiment had three situations, achieved by manipulating content concreteness (abstract, concrete and concrete that includes social contract). The participants' task was, as in the classical Wason task, to choose which cards need to be checked in order to test the validity of the displayed conditional rule. Types of answers, as well as response times were recorded. After every answer participants had to judge their confidence in the validity of their answer. The obtained results showed strong confirmation bias which was also accompanied with high metacognitive assessments as expected. Most of the participants in most of the tasks demonstrated confirmation bias by choosing the same two cards and that answer was false. At the same time they were very confident in their own answers. Surprisingly, concrete content did not decrease the bias effect. Obtained results are in line with the understanding that Type 1 thinking processes rapidly suggest the answer which is, in this case, false. But fluency of generating that answer affects the participant's metacognition and because of that she significantly overestimates her performance. Moreover, the judged confidence correlated with response times, which is in line with other research of metacognitive processes involved in reasoning tasks.

Keywords: *Wason selection task*, *metacognition*, *reasoning*, *judgements of confidence*

Introduction

The Wason selection task was developed by British psychologist Peter Wason (1966, 1968). This task has become one of the most used tasks in the psychology of reasoning. The usual example of this task goes like this:

> There are four cards on the table. Each card has a letter on one side and a number on the other side. The presented cards are A, K, 4, and 7. There is also a (conditional) rule which claims:
>
> If there is a vowel on the one side of the card, then there is an even number on the other side.

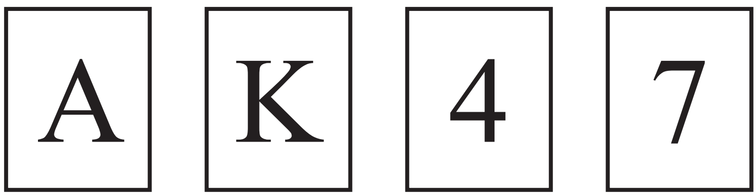

Figure 1. *An example of the Wason selection task with four cards*

The proposed rule could be correct or false. The participant's task is to select the cards that he (or she) thinks have to be turned over to test the truth of the rule.

The majority of the participants in the example above choose cards A and 4. And this answer is false. The only logically valid answer is selecting the cards A and 7. You can prove that by analysing the conditional rule "If [vowel] then [even number]" in the truth table.

Table 1. *Truth table for the conditional rule*

Vowel	Even number	If [vowel] then [even number]
yes	Yes	Yes
yes	No	No
no	Yes	Yes
no	No	Yes

As can be seen, the first card must be selected because it has a vowel and because of that it has to have an even number on the other side in order for the rule to be correct (first case in the table). If there is an odd number on the other side the rule must be false (second case in the table). That choice seems intuitive, and many subjects choose this card. The second card can have anything on the other side (third and fourth case in the table) so it is incorrect to choose it. The third card, which is often wrongly selected by participants, can also have anything on the other side (first and third case in the table) so it cannot be used to test the rule. And finally, the fourth card which presents an odd number *should not* have a vowel on the other side in order for rule to be correct (fourth case in the table). If there is a vowel on the other side then the rule must be wrong (second case in the table). So, the only way to test whether the rule is true or false is to select the first and fourth card presented in the example above. One can prove nothing about the rule by choosing the second and third card.

However the majority of participants give the wrong answers. Most of them select the first and third card. Many experiments have shown that participants give less than 10 % correct answers (Wason, 1966; Valerjev, 2000; Valerjev & Pedisić, 2002). Evans (1989) used the Wason task paradigm to explain biases in human thinking. According to his theory people are often biased in thinking processes. There are many such biases and the Wason selection task demonstrates at least two of them: the lexical matching bias (not considered here), and the confirmation bias. Evans explains the confirmation bias which has also been described as "the mother of all cognitive biases" (Dobelli, 2013). Basically, the confirmation bias is the tendency to search, interpret and remember information that confirms someone's preconceptions. Because of that, we do not perceive circumstances objectively and therefore it can lead to erroneous conclusions. It is manifested in many types of everyday thinking. In the Wason selection task confirmation bias is manifested in this way. Participants know that the rule might be right, so they select only the cards that seem to confirm the rule because they already have half of the confirmation (letter or the number that are related to the rule). Because of that they choose to check out the first and third card. The fourth card is not attractive according to confirmation bias, because it already contains the case that is not directly related to the rule. The participants use a shortcut in thinking and select only the cases that directly, positively test the rule, by proving it right, and not considering the cases that test the rule by using the negative examples. Participants only want to test the rule by confirming it and do not try to disconfirm the rule. This style of thinking is typical for most people. It can lead to serious mistakes, but on the other hand it is practical for everyday thinking in the real world when you have too many features that are somehow related and do not have enough time to separately check out each and every one of them. This type of thinking is quick and intuitive, it demands minimal effort and despite the fact that it can sometimes be misleading it provides us with decisions that are much better than random guessing. What we explained here is a typical description of a Type 1 process (or System 1 thinking) (Evans, 2009). In order for our thinking to be flexible enough in situations where System 1 fails, there is another way to obtain the conclusion. System 2 is slow, it requires mental effort and it is based on mental skills (like logical or mathematical skills) that usually have to be learned. In the example above, System 2 thinking would be active if the participant uses the rules of formal logic. When there is a feeling that something is wrong with intuitive answers, or when there is a conflict between more than one intuitive answer, the reasoner might be pushed from Type 1 to Type 2 thinking. The theory of thinking that explains these two processes is called dual process theory.

The topic of metacognition includes processes of monitoring and control over someone's cognitive performance. Traditionally, metacognition has been studied in the domain of memory. More recently, research on metacognition has spread into other domains of cognition, especially into the psychology of reasoning (Thompson, 2009; Thompson, Prowse Turner, & Pennycook, 2011). During the performance of different cognitive tasks people can monitor their own performance and make judgements about their efficiency and how hard or easy they perceive the task. There are many different measures of metacognition that depend on the time when they are taken (before, during or after the process of task solving), and on domain of cognition (memory, learning, reasoning, problem solving). In the psychology of reasoning the usual metacognitive measure is judgment of confidence. The participants have to estimate how confident they are that they solved the task correctly. The interesting question is how metacognition is correlated to actual reasoning efficiency. Studies in syllogistic reasoning have shown that reasoning accuracy and confidence are not correlated or loosely correlated but probably mediated by other processes (Shynkaruk & Thompson, 2006; Bajšanski, Močibob & Valerjev, 2014a; 2014b). On the other hand, significant negative correlations between response times and metacognitive judgements were obtained. Thompson et al. (2011) suggested that three types of cues can determine metacognitive estimates in reasoning tasks: answer fluency (the ease with which the initial conclusion comes to mind), conclusion acceptance and conflicting answers. Fluent answers, accepted answers and non-conflicting answers should be assigned higher confidence ratings.

It was already explained that the Wason selection task activates the Type 1 reasoning process which is fast but also results in a confirmation bias effect. However, use of the Wason selection task with concrete content sometimes can decrease the influence of the confirmation bias and increase the proportion of valid answers especially if it contains a social contract (Griggs and Cox, 1982; Stanovich & West, 1998; Valerjev, 2000; Valerjev & Pedisić, 2002). This change in answers is called the effect of thematic material. The thematic material effect can possibly be achieved by activation of another Type 1 reasoning heuristic, such as cheater-detection module, suggested by Tooby and Cosmides (1992). In that case, and if there is no conflict between the two or more answers generated by Type 1 processes, the new, correct answers would still be rapid and with high metacognitive judgment.

The aim of this research was to investigate metacognitive judgments in Wason selection tasks with different levels of content concreteness and the relation between metacognition and response time.

Materials, participants and procedure

Participants (N = 30) were recruited among undergraduate psychology students without prior knowledge of the Wason selection task. The experiment consisted of 30 trials divided into 3 groups: abstract material, neutral concrete material (arbitrary connections among concrete concepts) and social contract based material. Table 2 shows an example of each task.

Table 2. *Examples of trials according to material type*

		Card 1	Card 2	Card 3	Card 4
Abstract	If there is an even number on one side of the card, then there is a square on the other side.	4	5	Square	Triangle
Neutral concrete	If people are smiling, then it is sunny.	People are smiling	People are crying.	It is sunny	It is cloudy.
Social contract	If a person is drinking an alcoholic beverage, then they have to be an adult.	Drinking beer	Drinking soda	25 years old	16 years old

The experiment was designed in the E-Prime v2.0.10.356 software package and conducted on 5 identical PCs in the computer laboratory of the Department of Psychology. The 30 constructed trials were randomized for each participant as well as the order of the cards. Participants were instructed to decide, as quickly and as correctly as possible, which cards need to be turned over in order to validate the rule. After they chose which cards need to be checked they gave a metacognitive judgement of confidence for their choice on a continuous scale ranging from (0 % – not confident at all, I was guessing) to 100 % (completely confident in my answer) by typing the percentage. Figure 2 shows an example of how a single trial appeared on screen for the participants.

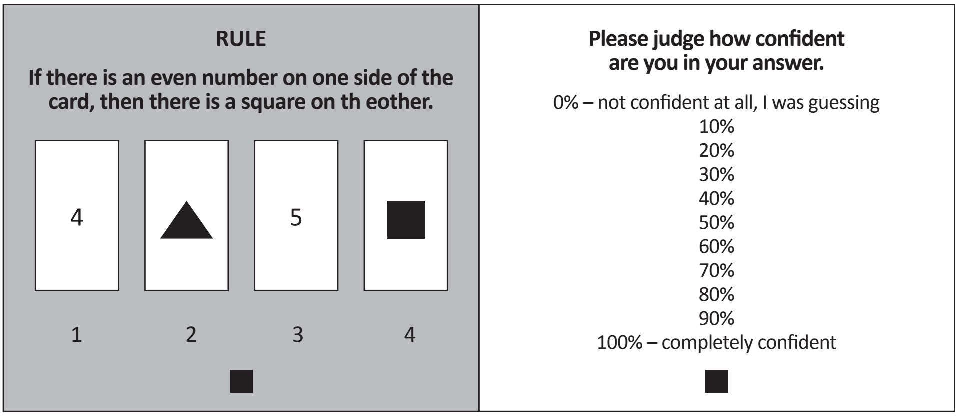

Figure 2. *An example of an abstract material trial as seen by a participant*

In the example shown the correct answer would be to turn over card number 1, and card number 2, in that case the participant would type 12 and hit the *ENTER* key (Figure 1, left). After choosing the cards, the metacognitive judgement scale appeared (Figure 1, right) and participants had to estimate their confidence by typing the percentage (participants were instructed they could type any number from 0 to 100 not just those shown here). Participants completed 3 practice trials and then proceeded to the main experiment. Choices, reaction times and judgements of confidence were recorded on each trial.

Results

Participants chose the correct combination of cards in a very small percentage of trials regardless of material type (4.67 % for abstract, 0 % for neutral concrete, and 1 % for social contract items).

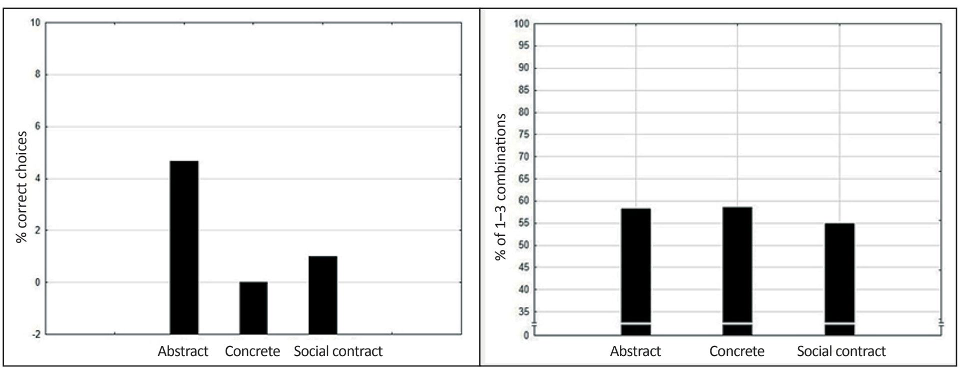

Figure 3. *Percentage of choices: correct (left), 1–3 combination (right)*

The classic confirmation bias response (1–3 combination; Figure 3, right) was chosen on a far greater percentage of trials (58.33 % for abstract, 58.67 % for neutral concrete, and 55 % for social contract items). These results are to be expected for abstract and neutral concrete material yet unexpected for the social contract trials which were expected to boost correct choice percentages.

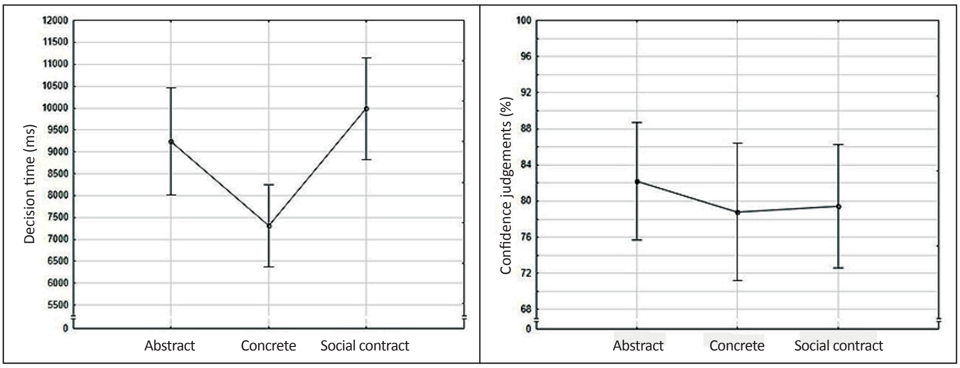

Figure 4. *Average decision times (left) and confidence judgements (right) for different types of trials*

Repeated measures analysis of variance showed decision times differed significantly depending on material type ($F(2, 58) = 12.27$, $p < .01$) with Tukey HSD post-hoc comparison showing decision times (Figure 4, left) for concrete trials were significantly smaller than for abstract and social contract trials. Judgements of confidence (Figure 4, right) were generally high (82.17 % for abstract, 78.81 % for neutral concrete, and 79.40 % for social contract items), but did not differ significantly ($F(2, 58) = 1.34$, $p > .05$) depending on material type.

To test whether slower decision making is connected to lower confidence judgements 3 Pearson correlation coefficients were calculated; one for each type of material. For abstract material the correlation approached significance ($r = -.36$, $p = .052$, $df = 28$) with slower participants giving lower confidence judgements. For both concrete ($r = -.43$, $p < .05$, $df = 28$) and social contract ($r = -.57$, $p < .01$, $df = 28$) trials the same correlation was significant. As has been shown in prior research slower participants generally show less confidence in their work.

Correlations connecting participant decision time and confidence judgements answer the question: “Are slower *participants* less confident in their choices?”, on the other hand, the same analysis conducted for items answers the question: “Are confidence judgements lower for *items* which participants processed for a longer time?”. The analysis is a Pearson correlation between average confidence judgements and decision times where items represent the sample. When conducting this type of analysis on all items the correlation does not approach significance ($r = -.26$, $p = .16$, $df = 28$), but when we removed neutral concrete items this changed drastically ($r = -.68$, $p = .001$, $df = 18$). It seems concrete items differ significantly from the other two types.

Discussion

Contrary to expectations, introducing social contract items did not boost accurate choice percentages. This might be explained by the specificity of the used method. Most studies use a smaller number of items and they are grouped into blocks, and mostly do not measure response times. In order to get a precise measure of response time more items were needed. Also, all items were randomly presented to participants rather than in blocks according to material type. This may have led to forming of a unique mental set for response selection more similar to classic Wason tasks which negated the influence of thematic materials. Emphasis on speed of decision might have also influenced participants when considering their choices. All of this may explain why both the percentages of accurate and biased responses were unified for all three experimental conditions.

Comparing decision times, it would seem concrete materials provided less processing load than abstract and social contract materials. There is a possibility that a matching bias (lexical matching between the conditional rule and cards) occurred for neutral concrete items, and not for the other two conditions. Since there was no lexical matching in the other two conditions participants needed an additional step to process the content which resulted in shorter decision times for neutral concrete items.

Judgments of confidence were high in all of the experimental conditions (Figure 4, right) even though objective accuracy was extremely low. In this particular task, participant reasoning was led by Type 1 processing, resulting in a large proportion of typical, rapid, but false choices. At the same time the fluency of Type 1 processing created the perception of high accuracy leading to extremely high confidence judgments. Participants tend to overestimate their performance. This is a typical finding of most studies regarding metacognition in memory (Benjamin et al., 1998; Koriat et al., 2004), learning (Hacker et al., 2000), and syllogistic reasoning (Shynkaruk & Thompson, 2006). Furthermore there were no significant differences in confidence judgments between experimental conditions. Possible lower processing load did not significantly increase metacognitive judgements for neutral concrete items. Judgments were already quite high for the other two conditions. It is possible that the additional processing in the other two conditions was not attributed to task difficulty, thus not contributing to judgments of confidence.

The result of slower participants giving lower metacognitive judgements was also apparent in this experiment, and present for all three material types (even if bordering significance for abstract trials). Item analyses showed different results for concrete items compared to the other two types. When considering only higher load items this analysis showed that decision times and confidence judgements were highly correlated. This finding is in line with prior research (Thompson et al., 2011). In total, results confirm the fluency-confidence connection while not making a distinction between the trial types when it comes to objective accuracy. Future research should concentrate on eliminating possible methodological problems. Presenting trials in blocks rather than randomly mixed trials of each type, changing the way in which participants respond, or streamlining the process would all refine the experiment. Further research is needed to determine why observed processing load differences did not lead to differences in confidence judgments.

Conclusion

As expected, participants overestimated their performance while objective accuracy was extremely low. Results also mainly showed a strong connection between fluency and confidence judgments, with further research needed to explain the findings for neutral concrete items. Introducing thematic material did not influence accuracy for the classic Wason task but the expected fluency-confidence connection was present. Item analysis showed a high negative correlation between decision times and confidence judgements for higher load items, but not for lower load items.

Acknowledgments

This work was supported by Grant 4139 from the Croatian Science Foundation.

References

Bajšanski, I., Močibob, M., Valerjev, P. (2014a). Metacognitive judgements and syllogistic reasoning. *Psychological Topics*, *23*, 1, 143–16.

Bajšanski, I., Močibob, M., Valerjev, P. (2014b). The effect of consensuality on metacognitive judgments in syllogistic reasoning. In P. Bernardis, C. Fantoni, W. Gerbino, (Eds.), *TSPC 2014 – Proceedings of the Trieste Symposium on Perception and Cognition*, (pp. 88–90) Trieste : EUT – Edizioni Università di Trieste.

Benjamin, A. S., Bjork, R. A., Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic Index. *Journal of Experimental Psychology: General*, *127*, 1, 55–68.

Dobelli, R. (2013): *The Art of Thinking Clearly*. New York: Harper Collins Publishers.

Evans, J. St. B. T. (1989). *Bias in Human Reasoning: Causes and Consequences*. Hove: LEA.

Evans, J. St. B. T. (1998). Matching Bias in Conditional Reasoning: Do We Understand it After 25 Years? *Thinking & Reasoning*, *4*, 1, 45–110.

Evans, J. St. B. T. (2009). How many dual-process theories do we need? One, two, or many? In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond*, (pp. 33–54). New York, NY, US: Oxford University Press.

Griggs, R. A. & Cox. J. R. (1982). The elusive thematic-material effect in Wason's selection task. *Quaterly Journal of Experimental Psychology*, *35*, 407–420.

Hacker, D. J., Bol, L., Horgan, D. D. & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92*, 1, 160–170.

Koriat, A., Bjork, R. A., Sheffer, L., Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*, 4, 643–656.

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

Stanovich, K. E. & West, R. F. (1998). Cognitive ability and variation in selection task performance. *Thinking and Reasoning*, *4*, 3, 193–230.

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, 107–140.

Tooby, J. & Cosmides, L. (1992). The psychological foundations of culture. In J. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*, (pp. 19–136). New York: Oxford University Press.

Valerjev, P. (2000). Wasonov izborni zadatak: Stabilnost efekata pristranosti i tematskog materijala. [Wason selection task: Stability of bias and thematic material effects]. *Radovi: Razdio filozofije, psihologije, sociologije i pedagogije*, *39*, 16, 97–111.

Valerjev, P. & Pedisić, A. (2002). Wasonov izborni zadatak – utjecaj upute, tipova kondicionala i tematskog materijala. [Wason Selection Task – Influence of Instruction, Conditional Types and Thematic Material]. *Radovi: Razdio filozofije, psihologije, sociologije i pedagogije*, *40*, 17, 45–64.

Wason, P. C. (1966). Reasoning. In B. M. Foss, (Ed.), *New Horizons in psychology 1*. Harmondsworth: Penguin.

Wason, P. C. (1968). Reasoning about the rule. *Quarterly Journal of Experimental Psychology*, *20*, 3, 273–281.