

# 10

## Metacognitive Monitoring in Wason Selection Task: The Influence of Content Abstractness, Conditional Type, and Social Content

**Pavle Valerjev**

Department of Psychology, University of Zadar  
valerjev@unizd.hr

### Abstract

*In meta-reasoning studies, solvers are asked to make a metacognitive judgment about the correctness of their reasoning. It is often found that solvers have no insight into the actual accuracy of their responses and that they make confidence judgments based on cues such as the fluency of the response. According to the dual-process theory of reasoning, responses to reasoning tasks generated by heuristic processes (Type 1) are automatic, quick, and easy, while responses generated by analytic processes (Type 2) are often slower, require mental effort, and are deliberate. Participants tend to have more confidence in their Type 1 responses than in their type 2 responses, regardless of their accuracy. The Wason selection task involves a form of conditional reasoning. Typical results show that accuracy in the abstract version of the task is very low. This can be explained by cognitive biases such as confirmation and matching bias, which are based on heuristic processes. The aim of this study was to investigate how variations in content abstractness, conditional type, and social content affect judgments of confidence in Wason tasks. 128 participants took part in the 2x2x2 experiment. They solved variants of Wason tasks and rated their confidence after each response on a 0-100% scale. The results showed a statistically significant effect of social content (the participants were more confident in non-social content tasks than in social content tasks) and abstractness (more confidence in abstract content tasks than in concrete tasks). It appears that heuristically generated responses promote higher metacognitive scores. This result is consistent with the general views of the meta-reasoning framework and dual-process theories of reasoning.*

**Keywords:** Wason selection task, metacognition, meta-reasoning, dual-process theory

## Introduction

Conditional reasoning is one of the most important areas of research in the psychology of reasoning. Researchers are interested in it because it involves a number of interesting effects that can be studied in tasks as simple as conditionals. The classical conditional reasoning task consists of two premises and a conclusion. The first premise is a conditional sentence with the form If P, then Q.

P is a simple proposition called the antecedent, and Q is a simple proposition called the consequent. The second premise and the conclusion are simple propositions that refer to the affirmation or negation of the antecedent and the consequent. This results in four possible forms of conditional tasks (see Table 1).

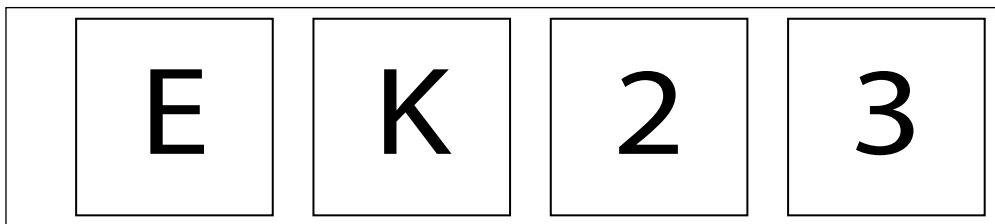
**Table 1**  
*Basic conditional task forms*

Name	Conditional premise	Second premise	Conclusion	Validity
Affirmation of antecedent (Modus ponens)	If P, then Q	P	Q	Valid
Affirmation of consequence	If P, then Q	Q	-	No valid conclusion follows
Negation of antecedent	If P, then Q	Not P	-	No valid conclusion follows
Negation of consequence (Modus tollens)	If P, then Q	Not Q	Not P	Valid

In addition to the simple conditional reasoning tasks with two premises, one of the most common conditional reasoning tasks is the Wason selection task (Wason, 1966, 1968). Alternative names for this problem are the Card selection task and the Four-card problem. This task has been studied for a long time and is an important tool in research on deductive and inductive reasoning, in research on scientific reasoning, and in research on cognitive biases. The classic version of the task (Wason, 1966, 1968) is an abstract version with a rule about the occurrence of letters and numbers on the cards. It reads as follows:

*TASK DESCRIPTION:*

*Each card in the set has a letter on one side and a number on the other. Four cards are chosen at random from the set and placed on the table. For example:*



*The experimenter then presents a hypothesis (conditional rule)*

*RULE:*

*If there is a vowel on one side of the card, then there is an even number on the other side of the card.*

*TASK FOR THE PARTICIPANT:*

*Select all the cards, and only those that need to be turned over, to find out if the hypothesis of the four cards on the table is valid or invalid.*

Note that according to the conditional rule *If P, then Q*, the four cards in the example above represent simple propositions P, not-P, Q, and not-Q. In this version of the task, the correct response is to select two cards: P and not-Q. For the rule to be valid, there must be an even number on the back of card E and no vowel on the back of card 3. The other two cards (not-P and Q) can have any letter or number on the back, so that they are not relevant to the validity of the rule.

However, the majority of participants select the incorrect combination of cards in this version of the task. They choose cards P and Q, but also other incorrect combinations. In the work of Wason and Johnson-Laird (1972), only 4% of participants chose the logically correct combination of cards. Among the incorrect responses, P Q was the most common, given in 46% of trials. When participants were differentially instructed to choose exactly two cards (and not any number of them), the dispersion of their responses decreased, but there was still an even stronger tendency to choose the typical incorrect response P Q (74%) over the correct response P not-Q (14%) (Valerjev, 2000; Valerjev & Pedisić, 2001). Initially, Wason (1968) explained this universal error as confirmation bias. Confirmation bias is the tendency to consider only the evidence and information that confirms our knowledge and beliefs and to ignore the evidence and information that does not confirm them. In other words, people tend to test their hypothesis only positively (verify it by choosing the P-card) and not try to test it negatively (falsify it by choosing the not-Q-card). Later, Evans (1984, 1989) attempted to attribute this effect to matching bias—the tendency to choose the cards that match the information contained in the question. Both biases are a typical result of heuristic reasoning processes that simplify the task and tend to speed up its solution. For a more detailed insight into the study of confirmation bias using card selection tasks and other reasoning tasks, see Evans (1989) and Valerjev (2024).

There are several variants of the Wason selection task, and some of them have led to a dramatic change in responses. One of the best-known is the manipulation through the content of the task. Instead of using abstract content, Griggs and Cox conducted a series of experiments using concrete (and social) content (see Griggs & Cox, 1982; Cox and Griggs, 1982). They used the conditional rule, which is more concrete, familiar, and social in nature: *If a person drinks beer, then the person must be over 19 years old*. The four cards presented (for the P, not-P, Q, and not-Q cases) were: *Drinks beer; Drinks Coke; 22 years old; 16 years old*. Participants' task was to select the cards that had to be turned over to determine whether someone was breaking the rule or not. The correct response was the same as the abstract version of the task: P and not-Q card (*Drinks beer* and *16 years old*). The change in content led to a clear shift towards correct responses. 60% of participants gave the correct responses (Cox and Griggs, 1982), and in the other version of the concrete task, where exactly two cards had to be selected, 56% of participants were correct (Valerjev, 2000; Valerjev & Pedisić, 2001). It was also found that participants tended to represent the abstract and neutral (not-social) conditional as biconditional, and that in this type of conditional the case P Q would also be the correct one.

The increase in accuracy has been described as the effect of thematic content (Wason & Shapiro, 1971). Although it was not always possible to replicate the effect when specific rules were used (see Pollard, 1981), this effect sparked a massive debate about which theory explained it better. The theory of pragmatic reasoning schemas (Cheng & Holyoak, 1985) assumes that when reasoning with concrete rules, we draw on learned rules (pragmatic schemas) from real life, which facilitates reasoning. The innate modules theory of reasoning (Cosmides, 1989) holds that we are sensitive to the violation of the social contract implicit in the rule because this sensitivity has been a useful evolutionary adaptation of humans. Mental model theory (Johnson-Laird & Byrne, 1991; Johnson-Laird, 2001; Johnson-Laird et al., 2015) states that we can use adequate mental models to generate the counterexample, but in the abstract version, the models may be incomplete and/or represent biconditional, which leads to an incorrect response. Formal rule theories of reasoning (e.g., Braine, 1978; Rips, 1983, 1994) have a problem explaining this

effect since both abstract and concrete rules have the same logical form. According to relevance theory (Sperber et al., 1995), we use the interpretation that is pragmatically relevant in the given context, which affects differences in responses. The probabilistic approach (Oaksford & Chatter, 1994, 2007) assumes that people draw conclusions that are rational in a probabilistic sense, i.e., that they choose answers that lead to the reduction of uncertainty. According to Evans' (1984, 1989, 2006) heuristic-analytic theory, the content and type of the conditional can influence the greater or lesser activation of heuristic or analytic thinking. Heuristic thinking leads to a typical incorrect answer, and analytic thinking leads to a correct one. For a more detailed comparison of the implications of the 16 theories of reasoning in the Wason selection task, see the analysis by Ragni et al. (2018).

For this study, we have adopted the heuristic-analytic approach proposed by Evans (2006), as well as other theories that fit into the dual-process approach to reasoning (for a detailed overview of developments in this area, see De Neys, 2018). The main idea is that there are two types of processes: Type 1, which tends to be fast and intuitive, more automatic and based on heuristics, and Type 2, which is slower and more deliberate, based on reflective and analytical thinking. Type 2 processes relate to the conflict in the responses and the resolution of the conflict by re-evaluating the response. Specifically, in the Wason selection task, this would mean that rapid heuristic processes (causing confirmation and matching bias or simplification of the conditional representation to a biconditional representation) lead to the typical incorrect response (P and Q). Manipulating the task material could increase the likelihood that the analytic processes are activated, which would lead to a higher proportion of alternative responses and, among these, to correct responses (P and not-Q).

Another important theoretical background for this study is the meta-reasoning framework by Ackerman and Thompson (2015, 2017). It provides explanations for the metacognitive monitoring and control that accompany processes of reasoning and problem-solving. According to this framework, people typically use a cue utilisation approach during reasoning. This means that they usually do not have access to the cognitive process itself (at least not to the automatic parts of it), which makes it difficult to assess how good the reasoning process and its responses are. Therefore, in metacognitive monitoring, people have to rely on indirect cues that are necessary for performing metacognitive control (e.g., decisions about strategy choice, further engagement in solving, etc.). The main cues that people typically rely on are fluency, ease and speed of responses, feeling that something is wrong, recognizing a conflict, duration of the task, etc. The cue utilisation approach has been observed in various reasoning tasks. For example, in the base-rate neglect task, the conjunction fallacy task, and the covariation detection task (see Dujmović & Valerjev, 2018; Dujmović, et al., 2020; Valerjev & Dujmović, 2019), metacognitive judgments were generally found not to correlate with actual reasoning accuracy (people often tend to overestimate their efficiency), but to correlate positively with fluency and negatively with the feeling of conflict. Participants are generally more confident when responses are faster and easier to produce and less confident when they are slower and more conflictual. In addition, responses produced by the dominant heuristic alone are more fluent and faster and are therefore judged with greater confidence in accuracy. In contrast, when there is a conflict between two or more intuitions or two or more processes, the final response is slower, generated with more effort, and therefore judged with less confidence.

A similar mechanism is expected in the Wason selection task. A typical incorrect response (P Q cards) results from heuristic intuition, which is quick and easy, and should, therefore, be associated with a high degree of confidence. The correct response (P not-Q) requires analytical thinking that involves rejecting the quick intuitive response and a deeper understanding of the conditional rule. Therefore, we expect that task variants with a high proportion of P Q responses will generally have higher confidence levels than tasks that elicit more non-heuristic response alternatives (including correct P not-Q responses).

A few studies have measured metacognitive parameters in the Wason selection task. In the work of Thompson et al. (2013), participants completed a modified abstract version of the Wason selection task in a two-response paradigm. The modification required participants to decide for each card whether it was useful in deciding the validity of the conditional rule and to rate their feeling of rightness (FOR) for their response. Each card was rated in two steps, once quickly and a second time giving participants enough time to re-evaluate their responses. The results showed that the responses consistent with the matching heuristic were faster, had higher FOR ratings, and were analysed less retrospectively. In another metacognitive study using a Wason selection task (Valerjev & Dujmović, 2017), participants were randomly presented with 30 Wason tasks (10 abstract, 10 concrete neutral, and 10 concrete with a social contract integrated into the conditional rule). Participants were instructed to respond as quickly as possible, which increased the strength of heuristic intuition, and resulted in participants relying almost exclusively on heuristic responding in all task variants, including the variant with a social contract. As a result, the proportion of correct responses was very low (less than 5% in all three task variants), and the proportion of incorrect heuristic P Q responses was high (55-58.67%). There was no difference between the proportions of the two response types in the different task variants. Confidence judgments were high (78.81-82.17%), and there was no difference in confidence judgments between task variants. The confidence ratings correlated with the response times so that the faster responses were generally rated with higher confidence.

Previous research has shown the expected effect of heuristic reasoning on metacognitive judgments in the Wason selection task. However, there is insufficient evidence for the change in metacognitive judgement levels associated with the shift in responses from heuristic to less heuristic reasoning in the Wason selection task. We expect that experimental manipulations in the Wason selection task that change responses from predominantly heuristic (P Q) to more analytic (P not-Q) and other non-heuristic responses would generally reduce confidence in such task variants.

## Aim

The aim of the study was to investigate changes in metacognitive monitoring in the Wason selection task. To this end, confidence judgments of response accuracy were measured in selection tasks that differed in terms of abstractness, conditional type, and social content.

## Method

*Participants.* The sample consisted of  $N = 128$  ( $N_{male} = 22$ ) psychology students from Zadar, Rijeka, and Mostar universities. The mean age of the participants was  $Age = 20.42$  ( $SD_{age} = 1.1$ ;  $min_{age} = 19$ ;  $max_{age} = 24$ ). The participants were rewarded with an appropriate research point.

*Research design.* The experiment was designed as a factorial  $2 \times 2 \times 2$  within-subjects experiment. Each of the three factors (abstractness, conditional type, social content) was manipulated by the two levels (abstract vs. concrete content; classical (non-deontic) conditional type vs. deontic conditional type; no social situation in content vs. social situation in content of the task). The experiment consisted of eight conditions. The recorded responses were selected cards and the confidence judgment for each response. Although the main interest was on the main factor effects, the balanced factorial design was chosen for better experimental control. By combining factor levels in the task variants, we wanted to capture broader task variants (e.g., the social effect was tested in both abstract and concrete tasks and in both deontic and non-deontic tasks).

*Materials and procedure.* Eight Wason selection tasks were constructed for this study. Each experimental condition was represented by one task. Table 2 lists the conditional rules used in the eight tasks. It was decided not to use more than one task for each condition to keep the procedure economical and simple. This was to avoid the undesirable uniformity of responses that occurs in experiments with too many tasks. Participants seem to give the same heuristic response for each task and ignore the experimental manipulation (see Valerjev & Dujmović, 2017) when there are many selection tasks and when there may be time pressure. The experiment was conducted in the classroom in several group sessions (before the lecture). The participants were presented with booklets with printed selection tasks. The order of the tasks in the booklets was rotated, so that the participants started with different tasks and had a different order. The participants were instructed to read the conditional rule carefully and then select the cards (out of four cards presented) to be flipped to determine whether the rule was valid or invalid. They then rated their confidence in the answer on a 0 to 100% scale. The zero value was anchored as “no confidence; I am just guessing”, while 100% was “complete confidence in the correctness of the response”. The participants then moved on to the next task. The procedure took about 10 minutes.

**Table 2**  
*Conditional rules that were used in eight variants of Wason selection task*

Experimental condition	Conditional rule to be tested in the task
A-nD-nS	If a card has a vowel on one side, then it has an odd number on the other side.
A-D-nS	If there is an even number on one side, then the other side of the card must have a capital letter.
K-nD-nS	If the weather is sunny, then the bees fly.
K-D-nS	If it's rainy, then the frogs have to croak.
A-nD-S	If a person has occupation X, then they are in cahoots with people from city ZZ.
A-D-S	If a woman has an occupation DDD, then she must be in contact with people from block C.
K-nD-S	If the athlete uses the sports hall, then they are a member of the “Athletics” club.
K-D-S	If a person drinks beer, then they must be of legal age.

*Note.* A = abstract and K = concrete, nD = not deontic conditional (classical conditional) and D = deontic conditional, nS = not social content and S = social content.

## Results

There are 15 ways to select four cards, resulting in 15 possible responses for each task. However, some of the combinations are pretty unlikely and uninteresting, given the typical responding patterns grounded in theories of reasoning. Two responses are essential for the aims of this study: P Q, which is very common and is an incorrect heuristic response resulting from heuristics that lead to confirmation and matching biases, and a correct P not-Q response that reflects understanding of the conditional rule and suppresses misleading heuristics and can therefore be considered an analytic response. All other responses are considered non-heuristic and were categorised as Other. Thus, there are three response categories: P Q, P not-Q, and Other.

**Table 3***Responses and judgments of confidence for eight variants of Wason selection task (N=128)*

Task	P Q(%)	Other (%)	P not-Q (%)	Conf. mean	Conf. SD	Conf. median	Skew.	Kurt.
AnDnS	93 (72.66)	27 (21.09)	8 (6.25)	88.68	20.25	100	-2.33	6.13
ADnS	89 (69.53)	32 (25.00)	7 (5.45)	86.27	23.45	100	-2.08	4.22
KnDnS	79 (61.72)	47 (36.72)	2 (1.56)	83.57	24.08	100	-1.78	3.12
KDnS	81 (63.28)	43 (33.59)	4 (3.13)	82.50	24.55	98.50	-1.61	2.35
AnDS	74 (57.81)	48 (37.50)	6 (4.69)	79.12	26.10	90	-1.26	0.97
ADS	74 (57.81)	47 (36.72)	7 (5.47)	77.90	28.43	90	-1.33	1.02
KnDS	64 (50)	56 (43.75)	8 (6.25)	75.56	28.43	80	-1.22	0.85
KDS	45 (35.16)	57 (44.53)	26 (20.31)	74.88	27.26	80	-1.20	0.98

*Note.* A = abstract and K = concrete, nD = not deontic conditional (classical conditional) and D = deontic conditional, nS = not social content and S = social content, Conf. = confidence judgment, Skew. = skewness, Kurt. = kurtosis.

The descriptive statistics of the results are shown in Table 3. It can be seen that the frequency of P Q responses, Other responses, and P not-Q responses changes in the different task variants. The highest frequency of P Q responses and the lowest frequency of Other and P not-Q responses were found for abstract/non-deontic/non-social task types. In contrast, the lowest frequency of P Q responses and the highest frequency of Others and P not-Q responses were found for the concrete/deontic/social task type. For the other task types, the frequencies for all three response categories were between the two stated frequencies.

The Cochran Q test showed that there was a statistically significant difference in the distribution of P Q responses across the eight variants of the task ( $Q = 89.22$ ,  $df = 7$ ,  $p < .001$ ), with the highest frequencies observed for the abstract/not deontic/not social variant. The same test showed a statistically significant difference in the distribution of valid P not-Q responses across the eight task variants, with the highest frequencies observed in the concrete/deontic/social task variant ( $Q = 62.22$ ,  $df = 7$ ,  $p < .001$ ) and in the distribution of all "Other" responses, which were more common in social content tasks than in not-social content tasks ( $Q = 41.13$ ,  $df = 7$ ,  $p < .001$ ). The frequencies of the different response types are shown in Table 3.

The mean values and medians of the confidence judgments for eight task variants are also shown in Table 3. It can be seen that the participants were generally very confident about the accuracy of their responses, while at the same time, their accuracy was quite low. This is a clear example of overconfidence. The highest confidence judgments are paired with responses of the abstract/non-deontic/non-social task type, and the lowest confidence judgments are paired with responses of the concrete/deontic/social task type. Many participants tended to rate their confidence very highly on most tasks, including those who gave the highest rating. For the first three task types, more than half of the participants rated their responses with 100% confidence in their accuracy (see the "Confidence median" column in Table 3). For this reason, the distributions of confidence values are negatively asymmetric, which can be seen in the "Skewness" column. Kolmogorov-Smirnov normality tests showed that all distributions deviated statistically significantly from the normal distribution (all  $p < .01$ ). Therefore, the non-parametric Wilcoxon matched-pairs test was chosen for the following analysis.

Three main factors with potential effects on confidence judgments were tested in the Wason rea-

soning task: abstractness, conditional type, and social content. Before the difference test, the confidence values of each task category had to be averaged for each participant (e.g., average of all non-social tasks vs. average of all social tasks). The average confidence values are shown in Table 4.

**Table 4**  
*Average confidence judgments in six task categories*

<b>Task category</b>	<b>Mean</b>	<b>SD</b>	<b>Median</b>
Abstract content	82.99	20.51	90
Concrete content	79.13	22.07	86.88
Non-deontic conditional	81.73	20.35	87.5
Deontic conditional	80.39	20.16	87.5
Non-social content	85.26	19.23	95
Social content	76.86	23.30	82.5

The Wilcoxon matched-pairs test showed a statistically significant difference in average confidence ratings for the responses in the Wason selection tasks about the social content factor. Note that the effect size  $r$  was calculated using Rosenthal's (1994) formula. Confidence ratings obtained in tasks with non-social and social content differed ( $Z = 5.63$ ;  $N = 128$ ;  $p < .001$ ;  $r = 0.5$ ). The participants were generally more confident in their responses to non-social content tasks than social content tasks. When comparing the abstractness of task content, the participants were generally more confident in their responses for tasks with abstract content than those with concrete content ( $Z = 2.86$ ;  $N = 128$ ;  $p = .004$ ;  $r = 0.25$ ). For the type of conditional rule in the Wason selection task (non-deontic vs. deontic), the difference in overall confidence in responses was not statistically significant ( $Z = 1.89$ ;  $N = 128$ ;  $p = .06$ ;  $r = 0.17$ ).

In addition, the confidence values of the heuristic P Q responses were compared with the confidence values of all other (non-heuristic) responses for all eight tasks. This grouping of responses was the only way to compare, as the frequency of specific responses (including the correct P not-Q response) was too low for most tasks. The non-parametric Mann-Whitney U test was chosen for the tests. There was a general tendency towards higher confidence ratings for P Q responses than for all other responses in all task variants. However, the difference was statistically significant in three task variants. For the concrete/not-deontic/not-social task ( $U = 1391.5$ ;  $Z = 2.66$ ;  $p = .008$ ;  $r = 0.24$ ), concrete/not-deontic/social task ( $U = 1468.5$ ;  $Z = 2.76$ ;  $p = .005$ ;  $r = 0.24$ ), and concrete/deontic/social task ( $U = 1386$ ;  $Z = 2.40$ ;  $p = .014$ ;  $r = 0.21$ ), the confidence ratings for P Q responses were higher than the confidence ratings for all other responses. The differences were not statistically significant for the other five task variants.

## Discussion

This study investigated whether the different variants of Wason selection tasks influence confidence judgments. The conditional rule manipulation included the social content factor, the abstractness factor, and the conditional type factor. The results showed that the participants were generally less confident in their responses in social content and concrete content tasks than in non-social content and abstract content tasks, while the conditional rule type (non-deontic vs. deontic) did not affect confidence. The participants tend to give fewer P Q responses (which are incorrect) in social content tasks and in concrete content tasks. In another words, the participants are more confident when inaccurate and less confident when they respond with non-heuristic responses (including the correct P not-Q response). At the specific task level, there is a tendency for the participants to make higher confidence judgments for P Q responses than for all other responses combined, at least for three task variants.

This finding is consistent with the general views of dual-process theories of reasoning (see De Neys, 2018; Evans, 2006), as well as with the explanations of the theoretical framework of meta-reasoning proposed by Ackerman and Thompson (2015, 2017). The heuristics activated in the Wason selection task led the reasoner to incorrect P Q responses that have been recognised as expressions of confirmation bias (Wason, 1968) and/or matching bias (Evans, 1984). There is overwhelming evidence that heuristic P Q responses are more frequent for abstract and non-social variants of the task than for concrete and social variants. This means that heuristic processes are stronger and more difficult to overcome in abstract and non-social variants of the task. On the other hand, it seems that concrete and social content in the task (for many different reasons—see the list of competing theories in Ragni et. al., 2018) promote overcoming the heuristic response and increase the likelihood of alternative, non-heuristic responses (including the correct P not-Q response). According to the meta-reasoning framework, metacognitive monitoring in reasoning is often based on cues rather than the actual accuracy of responses, and key cues include the fluency, ease, and speed of responses. Reasoners tend to assign higher confidence scores to heuristically generated responses because heuristically generated responses tend to be more fluent, easier, and faster than non-heuristically generated ones. Therefore, participants are generally more confident about incorrect responses when they are quick and easy, which is usually the case with heuristic responses. This finding is not uncommon and has been demonstrated in several other reasoning tasks, such as the base rate task (Dujmović & Valerjev, 2018) and the conjunction fallacy problem (Dujmović et. al, 2021), to name a few. It seems that there is a similar case with the Wason selection task.

This study has some limitations that need to be mentioned here. First, the study was not conducted in a laboratory but in classrooms using printed tasks, so it lacks rigour and control. The study also does not include measures of response times. Future studies should consist of response time measurement as it would be useful for comparing response times of different task variations. Finally, only one trial per experimental condition was used in this study (eight tasks for eight conditions), which opens up the possibility that some uncontrolled features of the tasks also influence the results. The reason for this decision was the fact that participants who have to solve too many Wason selection tasks would otherwise very soon start to give the same heuristic response (P Q) for all tasks (as shown in Valerjev & Dujmović, 2017). Even in the current study with a limited number of tasks, some participants gave the same P Q response and the same maximum confidence for all tasks. Therefore, the lesser of two evils was chosen and the participants were presented with limited tasks. To limit the number of tasks per participant, future studies could use a smaller number of main factors (perhaps only two) and/or independent samples of participants for different experimental conditions.

## Conclusion

Participants generally tend to give a high proportion of incorrect P Q responses in Wason selection tasks, which is probably a combined consequence of matching bias, confirmation bias, and the tendency to simplify conditionals and represent them as biconditionals. All of these reasoning styles belong to the variants of heuristic thinking. The participants were very confident in their responses in the tasks where they gave the highest proportion of incorrect (and heuristic thinking-generated) P Q responses. The manipulation of task content reduced both confidence scores and the frequency of P Q responses. Confidence decreased statistically significantly in tasks with social and concrete content compared to tasks with non-social and abstract content. Apparently, participants are most confident in the tasks where they are most likely to give quick, heuristic (and incorrect) P Q responses without considering the alternatives. Confidence in P Q responses is higher than that in all other responses in the three variants of the task. When the content of the tasks encourages other responses (including the correct P not-Q response), participants feel less confident in the response they choose. These findings are consistent with dual-process theories of reasoning and the meta-reasoning framework.

## References

- Ackerman, R., & Thompson, V. A. (2015). Meta-reasoning: What can we learn from meta-memory? In A. Feeney & V. A. Thompson (Eds.), *Reasoning as Memory* (pp. 164-182). Psychology Press.
- Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, 21(8), 607-617. <https://doi.org/10.1016/j.tics.2017.05.004>
- Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, 85(1), 1-21. <https://doi.org/10.1037/0033-295X.85.1.1>
- Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17, 391-416. [http://dx.doi.org/10.1016/0010-0285\(85\)90014-3](http://dx.doi.org/10.1016/0010-0285(85)90014-3)
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31, 187-276. [http://dx.doi.org/10.1016/0010-0277\(89\)90023-1](http://dx.doi.org/10.1016/0010-0277(89)90023-1)
- Cox, J. R., & Griggs, R. A. (1982). The effects of experience on performance in Wason's selection task. *Memory & Cognition*, 10(5), 496-502. <https://doi.org/10.3758/BF03197653>
- De Neys, W. (Ed.). (2018). *Dual Process Theory 2.0*. Routledge. <https://doi.org/10.4324/9781315204550>
- Dujmović, M., & Valerjev, P. (2018). The influence of conflict monitoring on meta-reasoning and response times in a base rate task. *Quarterly Journal of Experimental Psychology*, 71(12), 2548-2561. <https://doi.org/10.1177/1747021817746924>
- Dujmović, M., Valerjev, P., & Bajšanski, I. (2020). The role of representativeness in reasoning and metacognitive processes: an in-depth analysis of the Linda problem. *Thinking & Reasoning*, 27(2), 161-186. <https://doi.org/10.1080/13546783.2020.1746692>
- Evans, J. St. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, 75, 451-468. <http://dx.doi.org/10.1111/j.2044-8295.1984.tb01915.x>
- Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Hillsdale, NJ: Lawrence Erlbaum.
- Evans, J. St. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, 13, 378-395. <http://dx.doi.org/10.3758/BF03193858>
- Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, 73, 407-420. <http://dx.doi.org/10.1111/j.2044-8295.1982.tb01823.x>
- Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, 5(10), 434-442. [https://doi.org/10.1016/S1364-6613\(00\)01751-4](https://doi.org/10.1016/S1364-6613(00)01751-4)
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Lawrence Erlbaum.
- Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences*, 19, 201-214. <http://dx.doi.org/10.1016/j.tics.2015.02.006>
- Klauer, K. C., Stahl, C., & Erdfelder, E. (2007). The abstract selection task: New data and an almost comprehensive model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 680-703. <http://dx.doi.org/10.1037/0278-7393.33.4.680>
- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608-631. <http://dx.doi.org/10.1037/0033-295X.101.4.608>
- Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780198524496.001.0001>
- Pollard, P. (1981). The effect of thematic content on the 'Wason selection task'. *Current Psychological Research* 1, 21-29 (1981). <https://doi.org/10.1007/BF02684422>
- Ragni, M., Kola, I., & Johnson-Laird, P. N. (2018). On selecting evidence to test hypotheses: A theory of selection tasks. *Psychological Bulletin*, 144(8), 779-796. <https://doi.org/10.1037/bul0000146>

- Rips, L. J. (1983) Cognitive processes in propositional reasoning. *Psychological Review*, 90, 38-71. <https://doi.org/10.1037/0033-295X.90.1.38>
- Rips, L. J. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis*. (pp. 231-244). New York: Russell Sage Foundation.
- Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, 57, 31-95. [http://dx.doi.org/10.1016/0010-0277\(95\)00666-M](http://dx.doi.org/10.1016/0010-0277(95)00666-M)
- Thompson, V. A., Evans, J. St. B. T., & Campbell, J. I. D. (2013). Matching bias on the selection task: It's fast and feels good. *Thinking & Reasoning*, 19(3-4), 431-452. <https://doi.org/10.1080/13546783.2013.820220>
- 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. (2024). Pristranost potvrđivanja. [Confirmation bias]. In K. Damnjanović (Ed.), *Zbirka kognitivnih pristranosti i pogrešaka [Collection of cognitive biases and fallacies]*. (pp. 55-72) Heliks. <https://doi.org/10.31212/kogn.prist.2024.vale.03>
- Valerjev, P., & Pedisić, A. (2001). 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.
- Valerjev, P., & Dujmović, M. (2017). Metacognitive judgments during solving of Wason selection task. In N. Đapo, S. Zvizdić, A. Dautbegović, & M. Marković (Eds.), *Zbornik radova: Četvrti sarajevski dani psihologije*. (pp. 83-93). Filozofski fakultet Univerziteta u Sarajevu.
- Valerjev, P., & Dujmović, M. (2019). Performance and metacognition in scientific reasoning: The covariation detection task. *Psychological Topics*, 28(1), 93-113. <https://doi.org/10.31820/pt.28.1.5>
- Wason, P. C. (1966). Reasoning. In B. M. Foss, (Ed.), *New Horizons in psychology 1*. Penguin.
- Wason, P. C. (1968). Reasoning about the rule. *Quarterly Journal of Experimental Psychology*, 20, 3, 273-281. <https://doi.org/10.1080/14640746808400161>
- Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Harvard U. Press.
- Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *The Quarterly Journal of Experimental Psychology*, 23(1), 63-71. <https://doi.org/10.1080/00335557143000068>