IB Psychology
Internal Assessment

Title:

Investigating the effect of the nature of the Wason Selection Task on participants' performance.

Group: XYZ

Submitted: 10th October 2020

Word Count: 2169

Introduction

Thinking and decision-making is a fundamental cognitive process. One descriptive model of this process is the Dual-process theory, which outlines two systems of decision-making. System 1 is fast, unconscious thinking and automatic decision-making, used mostly for everyday decisions. It is based on past experiences and existing schemas, and uses heuristics to quickly make decisions. However, this comes at the cost of being error prone. System 2, on the other hand, is slow, conscious reasoning and controlled decision-making. It requires more concentration and cognitive resources. It is used for complex decision-making and is known to be more reliable than System 1 (Kahneman, 2013). Due to System 1's reliance on heuristics and pre-existing schemas, it is often more susceptible to cognitive biases. The Wason Selection Task (Wason, 1966) tested the confirmation bias as a consequence of System 1's automatic thinking. Participants were given a rule of the form *"if p, then q"*, and presented four cards showing *p*, *not p*, *q* and *not q*. They were asked to choose the cards that would verify whether the rule held or not. Participants tended to choose cards that confirmed the rule (*p* and *q*) rather than falsify it (*p* and *not q*).

Researchers **Griggs and Cox (1982)** argued that the Wason Selection Task was too abstract, and that participants' performance may differ if given a task with more relatable context. Hence, they conducted a study using the same procedure, but with the aim of investigating whether the nature of the task affected participants' performance. The sample consisted of 144 undergraduate students from the University of Florida. The research followed a repeated measures design; each participant was presented an abstract problem like the original Wason Selection Task, and realistic ones that had a more familiar context. As an example, one of the realistic problems used the rule: *"If a person is drinking beer, then the person must be over 19"*. It was found that participants performed significantly better when they could relate to the realistic task. Griggs and Cox proposed a memory-cueing hypothesis suggesting that performance on such tasks is facilitated when participants can recall past experiences within the context of the problem. In terms of the Dual-processing model, participants performed better on realistic tasks because System 1 thinking is based on past experiences and knowledge. The familiar context of realistic tasks allows System 1 to draw upon these experiences when performing the task, rather than resorting to heuristics or biases.

This study replicated the investigation of Griggs and Cox. The **aim** was to investigate whether the nature of the Wason Selection Task affected the performance of a sample of first year IB Psychology students. Some modifications were made to the original procedure, including a different realistic task; using two tasks in total; and having all participants perform the experiment together instead of in groups. The findings of this research can help facilitate the learning of students. If familiarity and context can help us better comprehend and solve problems, as well as improve System 1’s reliability, then this could be applied in pedagogy and to help students perform better on certain tests, such as Mathematics which involves questions that require similar critical thinking.

The **research hypothesis** states that the nature of the problem will have a significant influence on the participants' decision-making, which will be more accurate for realistic tasks. This will be

measured by the rate of correct responses on the tasks. The **null hypothesis** follows that the nature of the problem has no effect on the participants' decision-making. In other words, whether the task is realistic or abstract does not influence their performance. The **independent variable** in this investigation is the nature of the task, while the **dependent variable** is the participants' performance.

Exploration

This experiment followed a **repeated measures design**. This design was suitable for this investigation because it measures each participant's performance on both tasks, rather than the performance of a collective group on one task. This also helps control for participant variables, as each participant is subjected to all conditions.

The sample was collected using **convenience sampling**. This method was chosen because the investigation did not have strict exclusion criteria and it provided an efficient way of organizing a sample population. The sample consisted of 14 first year IB psychology students. Having only recently joined the class, they had not yet been exposed to either the original study or underlying theory. This made these participants a suitable choice for the investigation. The **exclusion criteria** consisted only of Year 2 psychology students, who were already aware of the original research and its findings.

Several **extraneous variables** were identified and controlled for during the investigation:

- The original realistic task was modified to be more relevant to the sample population.
- Participants were given a time limit for each task, as they may switch to System 2 if given too much time.
- Answers were collected at the end of each task, to ensure participants did not alter their answers to the previous task.

Materials used included:

- Abstract task (*"If a card shows an even number on one face, then its opposite face is blue"*).
- Realistic task based on the Covid-19 pandemic (*"If you go outside, you must wear a mask"*). This was chosen over the original drinking age problem because it would be more relevant to the sample, which was below drinking age.
- Image projector. One task was projected at a time, for a limited time, to ensure participants did not switch to System 2 and focused on one task only.
- Answer sheets (One per task per participant). These were collected afterwards to ensure answers could not be altered.

The **procedure** was as follows:

1. Participants were gathered in one room.
2. Consent forms and answer sheets were distributed (See Appendices I & III).
3. Participants were briefed. Standardized instructions and their rights were read out (See Appendix II).
4. Consent forms were signed and collected.
5. First task was displayed for two minutes.
6. Answer sheets were collected, and new ones distributed.
7. Second task was displayed for two minutes.
8. Answer sheets were collected.

9. Participants were debriefed (See Appendix VIII), and their questions were answered.

Analysis

	Abstract	Realistic
Mode	B, C (5)	A, C (6)
Variation Ratio	0.643	0.571
Frequency of correct answers	3 (B, D)	3 (B, C)

Table 1: *Descriptive Statistics.*

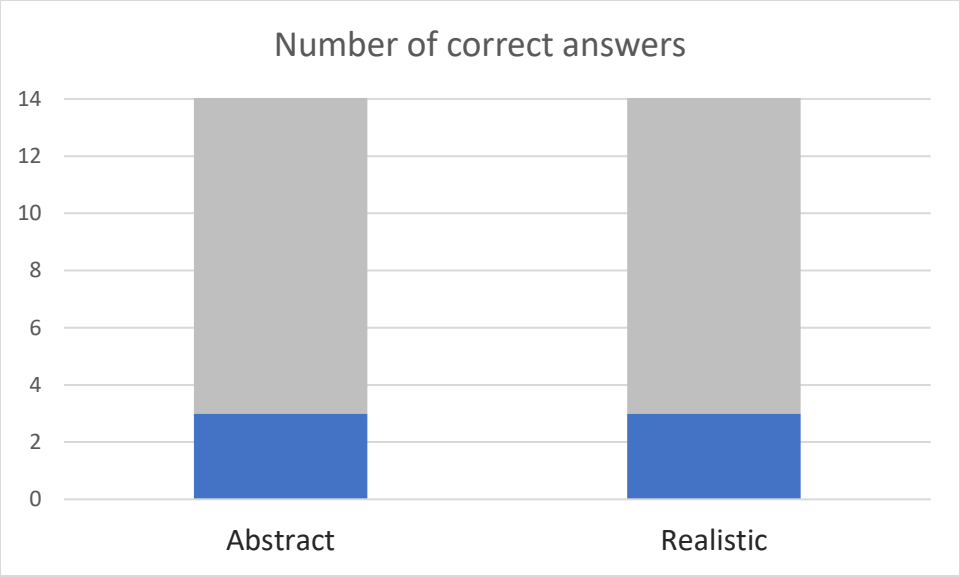


Figure 1: *Comparison of correct answers for abstract task and realistic task*

Table 1 and **Figure 1** represent the mode, variation ratio (VR) and frequency of correct answers for the data collected (See Appendix V for raw data). The data obtained was nominal, with each category represented by the answers chosen. As measures such as mean and median for central tendency or standard deviation for dispersion are not possible with nominal data, mode was chosen as the most suitable measure of central tendency, and VR (VR = $1 - \frac{mode}{sample\ size}$) for dispersion. As shown in **Table 1**, the mode for the abstract task was "B, C", and "A, C" for the realistic. Under the rule *"if p, then q"*, both would be equivalent to the answer "*p*, *q*". (See Appendix IV for how the tasks corresponded to the *"if p, then q"* rule). In other words, the data shows that for both tasks, participants tended to choose the option that *confirmed* the rule, rather than falsify it. This agrees with the findings of the original Wason Selection Task.

However, as **Figure 1** shows, the frequency of correct answers was the same for both tasks, indicating that there was no difference between overall performance on the two tasks. Despite this, the VR for the realistic task was lower, indicating that answers for the realistic task were less dispersed. The frequency table (See Appendix VI) also showed that for the realistic task, all answers from participants fell under three categories, whereas for the abstract task the answers fell under five categories. This suggests that although the number of correct answers was the

same for both tasks, there was still a pattern in the participants' thinking and comprehension of the realistic task, whereas for the abstract task it was more random.

A binomial sign test was used for statistical significance, chosen as the most suitable for this experiment due to the nominal data and repeated measures design. For each participant, an improvement in performance on the task was recorded as positive, while a decrease was recorded as negative. Out of 14, 12 participants' performance was invariant on both tasks. These results were omitted, while the remaining two were used for the binomial sign test. However, this number was below the minimum needed for the test, indicating that the data can be considered statistically insignificant, and the research hypothesis was rejected (See Appendix VII for the binomial sign test). Based on this, it can be said that the nature of the task has no effect on the sample's performance. In terms of the dual-processing model, these findings disagree with the memory-cueing hypothesis, suggesting that past experience does not significantly facilitate System 1 decision-making in the current sample.

Evaluation

As the results have shown, the research hypothesis was rejected. The findings of this investigation did not support those of Griggs & Cox, who found a significant improvement in the number of correct answers on the realistic task whenever participants had prior experience or familiarity with the context of the task. Moreover, the results of this investigation do not support the memory-cueing hypothesis either, which proposes that System 1 thinking is not as error prone as the Dual-processing model suggests when participants are able to rely on past experience.

One strength of using a repeated measures design was low participant variability, as each participant was subjected to both the conditions of the experiment. This minimized the effect of individual differences on the tasks, such as problem solving or language comprehension skills. However, this design also introduced order effects and demand characteristics which may have affected the validity of the data collected, as all participants performed the tasks in the same order. In future investigations, counterbalancing can be used to minimize order effects by splitting the sample into two groups and having them perform the tasks in different orders.

The convenience sampling used allowed gathering participants efficiently, and the exclusion criteria improved the validity of the responses since participants did not have prior knowledge of the experiment or the task. However, certain characteristics of the sample may have influenced the results. One such characteristic was that all participants except one were second language learners of English. This may have affected their ability to fully understand the instructions – which were delivered verbally – or the tasks, both of which were presented in English. In the future, written instructions and translations can be provided to ensure participants understand the procedures.

One strength of the study came from the usage of the projector for conducting the tasks. Displaying the tasks on the projector allowed to accurately limit and control the time spent on both tasks, thus eliminating an extraneous variable, and improving reliability. However, the way tasks were displayed to participants could have affected their understanding of the experiment. Each task slide displayed the rule for the task, and the four possible options for the rule, shown on cards corresponding to each option (for example, a red card, or a card showing ‘5’). However, the cards were then also labeled from A to D, and participants were given answer sheets with the options A to D and asked to circle their answer. This may have caused confusion among the participants, as the options provided on the answer sheets did not directly correspond with the information shown on the cards. Next time, it would be better if the answer sheets displayed the same options as the cards, rather than assigning letters.

Variables beyond control may have also affected the outcome. For instance, it was not possible to know whether participants actually used System 1 or System 2. When told they would be answering questions, some participants may have directly used System 2. On the other hand, participants may have attempted the first task with System 1, but then switched to System 2 for the second, since both tasks were similar and required complex decision-making. In the latter

case, a filler task could help to account for it to some degree, however whether participants used System 1 or 2 would still be largely out of the control of researchers.

From the results of this study, we may conclude that the nature of the Wason Selection Task does not influence participants' performance. For future investigations, it would be interesting to test if the *amount* of past experience with real-life situations affects performance on the task. This investigation used the Covid-19 pandemic as the real-life situation, which would still be quite a recent experience for participants. Such an investigation would give more insight into the memory-cueing hypothesis, as well as how participants can be trained to become better at these problems by giving them more experience with the context of the task.

Works Cited

*Wason, P. C. (1966). "Reasoning". In Foss, B. M. (ed.). New horizons in psychology. **1**. Harmondsworth: Penguin. LCCN 66005291.*

Kahneman, D. (2013). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology, 73*(3), 407-420. doi:10.1111/j.2044-8295.1982.tb01823.x

Appendix I: Consent Form

Hello everyone,

We, DP2 psychology students are currently carrying out an experiment as a part of our Internal Assessment to fulfill our IB Psychology class requirements. To take part in this experiment, we would like you to have more knowledge of your rights. As a participant, you will have to sign the consent letter enouncing your willingness to join the experiment. If you agree to participate in this experiment, you should know that:

- You should be at least 16 years old.
- You understand that you are participating in a psychological research/study.
- You have not experienced any major accidents.
- You may withdraw from the experiment at any time as well as withdraw your data after results are obtained.
- All data obtained throughout the experiment will be kept confidential and remain anonymous.
- You will be informed about the nature of the study and may receive the findings of the study after researchers have obtained the results.
- You are not allowed to bring any snacks or drinks into the study.
- Phones need to be placed in the pockets placed at the back of the classroom.

The experiment will take around 20-25 minutes to complete.

Date: Monday, August 24, 2020

Time: 3:30 PM

Location: 4th floor, Psychology classroom

We appreciate your participation in the experiment and are grateful for your support!

CONSENT

If you choose to take part in the experiment, we would like you to ask to sign the form below and fill in the information regarding the experiment.

I, _____, understand the nature of the experiment and I agree to participate in it voluntarily. I give the permission to the DP2 psychology students to make use of my personal data as a part of their experiment.

Age: _____ Gender: _____ Mother tongue: _____

Signature: _____ Date: _____

Appendix II: Standardized Instructions

Welcome to our psychology experiment.

1. We are now conducting a psychology experiment for our psychology internal assessment. We have the permission from the supervisor to conduct this experiment.
2. We will keep all your information confidential, and none of your personal data will be made public.
3. You have the right to request that your data be withdrawn from the experiment if you wish.
4. You have the right to request the final report and findings of this experiment, when available.
5. You are not allowed to leave the experiment halfway. If you want to leave, please leave now.
6. Please put away your phones in the pocket. No materials are needed other than a pen/pencil, and the consent form and answer sheets which will be provided to you.
7. Please do not talk to each other during the experiment.
8. For further information and instructions, you need to sign the consent letter.
9. The consent form is on the left side of the table, please read it carefully and fill out the form.

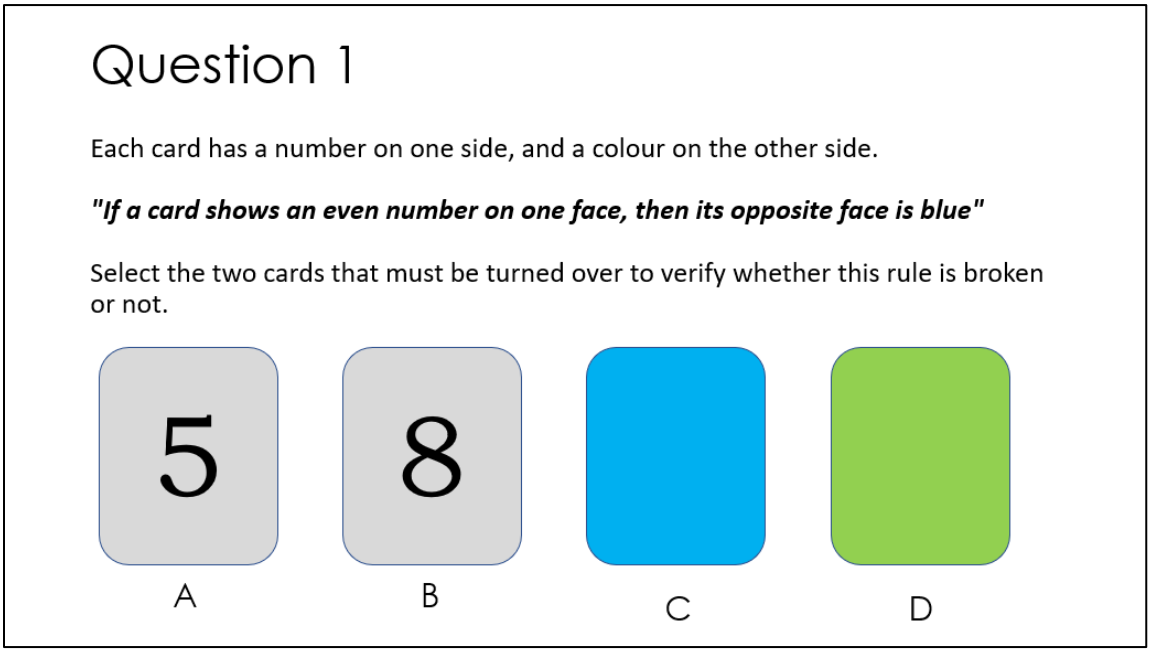
(After collecting the forms)

1. During the experiment, please follow instructions carefully. If you have any questions, you can raise your hand and we will come to you. However, please keep in mind that we will not answer any questions relating to our experiment's purpose until the end of the procedure.
2. In the experiment, you will be required to:
    a. Ensure that you have a pencil/pen and have been given an answer sheet. No other materials are allowed in the experiment.
    b. You do not need to write down your name or grade on the answer sheet. Simply circle the options provided to indicate your answer.
    c. Read through the questions presented on the projector carefully, and mark down your answer on the answer sheet. There are two questions in total and will be presented one by one. You will have 120 seconds or 2 minutes to answer each question.
    d. The answer sheets will be collected after each question.
    e. After all the questions have been answered, please remain seated patiently for the debriefing session.
    f. Are there any questions?

Appendix III: Answer Sheet

Appendix IV: Tasks

Abstract Task Slide:

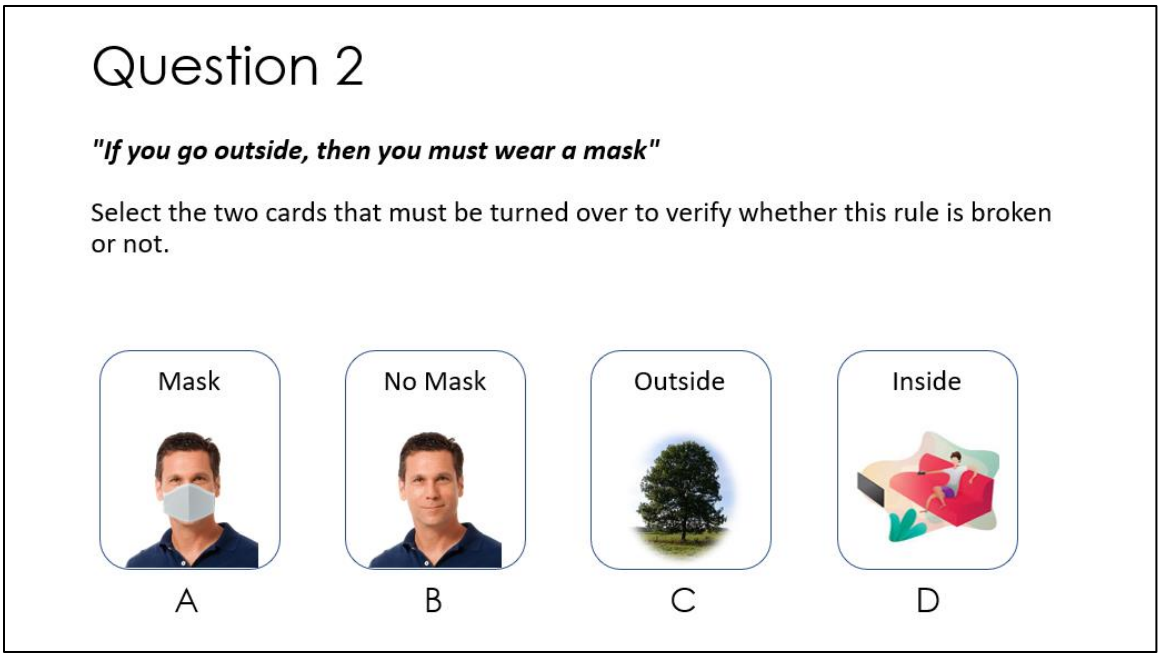


$$p = Even\ Number, \quad q = Blue\ Face$$

$$A \rightarrow not\ p, \qquad B \rightarrow p, \qquad C \rightarrow q, \qquad D \rightarrow not\ q$$

The choice of **B** and **C** confirms the rule "if p, then q", but does not verify it. The correct answer is **B** and **D.**

Realistic Task Slide:



$$p = Going\ outside, \qquad q = Wearing\ a\ mask$$

$$A \rightarrow q, \qquad B \rightarrow not\ q, \qquad C \rightarrow p, \qquad D \rightarrow not\ p$$

The choice of **C** and **A** confirms the rule "if p, then q", but does not verify it. The correct answer is **C** and **B**.

Appendix V: Raw Data

ID	Abstract	Realistic
1	A, C	B, C
2	C	A, C
3	B, C	A, C
4	A, D	A, C
5	B, C	B, D
6	A, C	B, D
7	A, C	A, C
8	B, C	A, C
9	B, D	A, C
10	B, D	B, C
11	B, C	B, D
12	B, D	B, C
13	A, C	B, D
14	B, C	B, D

Appendix VI: Frequency Table

Abstract		Realistic	
Answer	*Frequency*	*Answer*	*Frequency*
B, C	5	A, C	6
A, C	4	B, D	5
B, D	3	B, C	3
A, D	1	-	-
C	1	-	-

(Correct answers highlighted in green)

Appendix VII: Inferential Statistics

Binomial Sign Test:

ID	Abstract	Realistic	Difference
1	0	1	+
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	1	0	-
10	1	1	0
11	0	0	0
12	1	1	0
13	0	0	0
14	0	0	0

To obtain an adjusted sample size of *n* for performing the binomial sign test, wrong answers were marked as 0 and correct as 1. An improvement in performance (0 on abstract, 1 on realistic) was marked as positive. A decrease in performance (1 on abstract, 0 on realistic) was marked as negative. Counting all the positives and negatives gives $n = 2$.

n(+)	1
n(-)	1
n	2

The following table was then consulted to test for significance:

n	0.05	0.01
8	1	0
9	1	1
10	1	1
11	2	1
12	2	2
13	3	2
14	3	2
15	3	3
16	4	3
17	4	4
18	5	4
19	5	4
20	5	5

If n(+) or n(-) is below the critical values given for n, the data can be considered statistically significant. However, the minimum value of n needed is 8. Since the value of n for this investigation was below the minimum, the data was not significant, and the null hypothesis was retained.

Appendix VIII: Debriefing Note

The experiment is now over. We want to thank you for your corporation. The true aim of this study was to replicate the original research study conducted by Griggs and Cox in 1982, which itself was an extension of the Wason Selection Task. You may have noticed that the two questions presented were quite similar. This type of question is known as the Wason Selection Task, and is often used in studies on deductive reasoning and decision-making. The aim of the study conducted by Griggs and Cox was to investigate if the nature of the problem would influence the participant's performance on the test. The nature of the problem refers to whether the task is more abstract or more realistic. You may have noticed that this was the difference between the two problems presented to you today. The abstract one used numbers and colors, while the realistic one used the current real-life situation that we are all in, i.e having to wear masks when going outside due to covid-19. The original research used a different problem for the realistic task, which we modified so that it would be more relatable to our participants. The original study found that people performed significantly better on the realistic task than the abstract one. And this is important for research in psychology, because a lot of studies on thinking and decision-making are based on abstract tasks performed by participants. These studies do not take into account the impact of context on the thinking and decision-making abilities of humans.

Are there any questions?

All the data you provide will be kept confidential. If you would like to know the results of the experiment or read the final reports, you are welcome to do so at the completion of them. If you have any further questions, please do not hesitate to ask.

Now you may leave.