

Wason Selection Task and Variations

Nicholas Ciuica

Department of Cognitive Science, RPI

COGS 2120: Introduction to Cognitive Science

Dr. Bram van Heuveln and Dr. Ben Weissman

April 21, 2025

Wason Selection Task and Variations

Included here is a link to the GitHub project accompanying this research (Ciuica, 2025/2025):

<https://github.com/NicholasCiuica/WasonTask>

The Wason Selection Task (WST) is a type of selection task in which participants are shown four cards and asked to select which cards must be flipped to test the validity of a proposed rule. This rule is in the form of a conditional proposition expressed as “if P, then Q,” and each of the four revealed faces of the cards shows information aligning with P, Q, not-P, and not-Q (Nakamura, n.d.). The original WST was performed using cards that had a letter on one side and a number on the other. The task’s rule had two variations, one that read “If there is a D on one side of any card, then there is a 3 on its other side,” and the other being a reversal of the first rule’s antecedent and consequent. Wason also ran experiments with cards that had shapes on one side and colored scribbles on the other (Wason, 1968). There are four types of conditional propositions given an antecedent P and its consequent Q: modus ponens (MP), P implies Q; denying the antecedent (DA), not P implies not-Q; affirming the consequent (AC), Q implies P; and modus tollens (MT), not-Q implies not-P. Only MP and MT are valid forms of conditional reasoning. The goal of the WST experiment was to determine whether participants would be able to discern the correct pair of cards that align with these propositions (Zhang et al., 2021). Since Wason’s initial experiments, many WST studies have been run to explore the effects of different rules and card contents on participant’s responses, and to attempt to explain why people often make mistakes in selection tasks by designing models and algorithms (Ragni, n.d.).

The first WST experiments resulted in four canonical card selections, P, PQ, PQ not-Q, and P not-Q, and each of the other twelve possible selections had a response frequency less than

chance (Ragni et al., 2018). Many theories have been proposed to explain the occurrence and frequencies of these canonical selections, but only two models have withstood criticism for a number of reasons, including the fact that many other models work under the flawed assumption that the selection of each card in the task is independent (Ragni, n.d.).

The first of these models is the insight model, which was designed by Jonhson-Laird and Wason (1970) and implemented into an algorithm. This model categorized participants of a selection task by their level of insight in falsification thinking. To properly test the validity of a selection task's rule, participants must disprove counterexamples that could falsify the rule, rather than trying to prove the conditional statement by finding examples that verify the rule (Zhang et al., 2021). Three levels of insight are defined: Those with no insight will only select cards mentioned in the rule (selecting P or PQ, known as matching/verification bias), using MP and AC logic in an attempt to verify the rule; those with partial insight consider all cards, and in addition to selecting cards that verify the rule, they may also select cards that falsify the rule, including not-Q in their selections of P and Q; Those with complete insight know that they only need to select cards that can falsify the rule to test its validity, and will select P not-Q (Jonhson-Laird et al., 1970). This model explains the common occurrence of incorrect responses and matching bias as a lack of insight in the majority of participants of selection tasks, and this theory also implies that participants need previous training in falsification thinking to reach complete insight (Zhang et al., 2021).

The second model is the inference-guessing model, which is composed of ten parameters describing the reasoning steps participants take in a selection task (Klauer et al., 2007). These parameters have values within a probability interval of 0 to 1, and they are used to represent, for example: the probability that a participant guesses instead of making a logical inference; the

probability that a certain card is selected when guessing using heuristics (e.g. matching) rather than inference; the probability that “if P then Q” is interpreted as a conditional statement, rather than a biconditional one; and so on. A notable flaw of the original inference-guessing model is that it does not account for the selection of more than two cards through inference.

The insight model is currently the preferred selection task model because it better fits experimental data (Ragni, n.d.). It is also simpler than other models, like the inference-guessing model, that use parameters to account for guessing or for non-canonical outcomes with probabilities less than chance. Additionally, the insight model is able to explain the effects of certain factors (like previous knowledge and variations of rule and card contents) on participants’ ability to make the correct selection as an increase in the participants’ insight into falsification thinking. Iterations of the WST experiment have been categorized as abstract, everyday, or deontic based on the content with which they test these factors.

Abstract variations of WST are those that do not have cards or rules which participants may have had previous associations or everyday experiences with (Ragni, n.d.). This includes Wason’s original experiments and variations that make associations between abstract content like symbols or colors. Across time and space, participants consistently make mistakes in abstract WST, with multiple studies reporting less than 10% of participants choosing the correct pair of cards (Zhang et al., 2021), and a meta-analysis reporting that only 19% of participants make the correct selection in abstract WST studies (Ragni, n.d.).

The same meta-analysis also noted that participants of WST studies with everyday contents (information that is familiar to subjects, also called “concrete”) pick the correct pair of cards 29% of the time, implying that participants’ previous knowledge and experiences may assist them in engaging in falsification thinking (Ragni, n.d.). One example of an everyday task

is Griggs and Cox's transit task (1982) (although they called it a "thematic" task); undergraduates from the University of Florida were shown four cards with revealed faces labeled "Miami", "Atlanta", "plane", and "car" and then presented the rule "Every time I go to Miami I travel by car" or any of the other four rules generated by substituting in one of two city names and modes of transport (this task's rule uses "every" in "universal form" as opposed to a traditional "if-then" statement). It was believed that these Floridian students would be familiar enough with the kind of transportation they'd take to get to the locations shown on the cards that they would do better on the selection task. As a control, participants were split into two groups that would either participate in an abstract task before or after the transit task. When comparing the abstract and transit tasks, although the correct response rate was calculated to be 9% for the transit task and less than 1% for the control abstract task (no correct answers were given), the difference in performance between the tasks was not deemed statistically significant. Griggs and Cox corroborate their findings with a number of other "thematic" studies: two previous studies also found no effect in adding everyday content, two studies found a significant but weak effect, and two more found a significant effect with only some of the everyday contents that they tested (Griggs, 1982). The inconsistent results of everyday WST studies could be due to overestimation of the strength of associations that participants have with familiar material. For example, while the students from Florida in the transit task are familiar with Miami, Atlanta, planes, and cars, any associations they have with these concepts may not match with the task's randomized rule.

Some experiments have also focused on the effects of "deontic" (social/ethical) contents within WST. Participants have been shown to perform significantly better on selection tasks with deontic rules than those with abstract or everyday rules (Ragni, n.d.). In addition to their everyday WST, Griggs and Cox (1982) also ran an experiment on a deontic task involving legal

drinking age (which at the time of the experiment was 19): undergraduates age 18-22 from the University of Florida were shown four cards with the phrases “drinking a beer”, “drinking a coke”, “16 years of age”, and “22 years of age”, and were presented with the rule “If a person is drinking beer, then the person must be over 19 years of age.” Participants were asked to imagine themselves as a police officer looking for cases of illegal drinking. The authors noted that this experiment provided evidence for Manktelow & Evans’s memory-cueing hypothesis, which supports the idea that participants would have better performance on selection tasks if the contents of the task cued them to previous experiences. This task was also paired with a control abstract task, and this time there was a significant difference in performance; 74% of participants correctly responded to the deontic task, while nobody correctly responded to the abstract task (Griggs, 1982). This finding is similar to the 69% correct response rate reported by the meta-analysis (Ragni, n.d.). An important note about this substantial improvement in performance is that there was no observed “transfer effect” (e.i. subjects that did the deontic task first didn’t do any better on the abstract task), which is further evidence for the memory-cueing hypothesis (Griggs, 1982).

Selection tasks like WST have shown that reasoning within the brain does not necessarily follow the rules of formal logic, but often uses heuristics and biases to simplify tasks. This can lead to participants frequently making mistakes in the context of abstract logical thinking. However, research suggests that human brains may have evolved a cheater-detection mechanism that allows for increased performance in social “detective” roles. This is made evident by the difference in performance between abstract or everyday tasks and deontic ones (Van Lier et al., 2013). While more research is needed to improve our understanding of logic in the human brain (e.g. how does intelligence affect selection tasks responses?, what causes participants to make

errors or guess on logical tasks?), the insight model has proven useful in its ability to explain the role of task content in participants' logical abilities (Ragni, n.d.).

References

- Ciuica, N. (2025). *NicholasCiuica/WasonTask* [JavaScript].
<https://github.com/NicholasCiuica/WasonTask> (Original work published 2025)
- 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://cogsci.bme.hu/~ktkuser/KURZUSOK/BMETE47MN05/2017_2018_2/Griggs_Cox_1982_The-elusive-thematic-materials-effect-in-Wasons-selection-task.pdf
- Johnson-Laird, P. N., & Wason, P. C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, 1, 134-148.
- 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.
- Nakamura, K. (n.d.). *Similarities and Individual Differences in the Wason Selection Task: An Item Response Theory Analysis*.
https://escholarship.org/content/qt2704w91r/qt2704w91r_noSplash_9cd50904d4bac8f5741ba3969cbd84b1.pdf?t=op2jme
- Ragni, M. (n.d.). *The Wason Selection Task: A Meta-Analysis*.
https://escholarship.org/content/qt3rv0k45d/qt3rv0k45d_noSplash_8646c09c2d9744d4cabdbda295edd889.pdf?t=sgiit1
- 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://www.modeltheory.org/papers/2018selection.pdf>

- Van Lier, J., Revlin, R., & De Neys, W. (2013). Detecting Cheaters without Thinking: Testing the Automaticity of the Cheater Detection Module. *PLoS ONE*, 8(1), e53827. <https://doi.org/10.1371/journal.pone.0053827>
- Wason, P. C. (1968). Reasoning about a Rule. *Quarterly Journal of Experimental Psychology*, 20(3), 273–281. <https://doi.org/10.1080/14640746808400161>
- Zhang, M., Wang, L., Zou, F., Wang, Y., & Wu, X. (2021). The Brain Structure and Intrinsic Characters of Falsification Thinking in Conditional Proposition Testing. *Frontiers in Human Neuroscience*, 15, 684470. <https://doi.org/10.3389/fnhum.2021.684470>