Wason's Cards: What is Wrong?

Pei Wang
Computer and Information Sciences, Temple University
<peiwang@mindspring.com; http://www.cogsci.indiana.edu/farg/peiwang/>

This paper proposes a new interpretation of Wason's selection task. According to it, the results of the experiment do not show that human reasoning is not logical, but that the traditional logic is not a proper normative theory of reasoning under certain conditions. A new logic is introduced, which is consistent with the experiment results of the task.

THE SELECTION TASK

In Wason's selection task, subjects see four cards showing symbols like ***E***, ***K***, ***4***, and ***7***, and are told that each card has a letter on one side and a number on the other. The task is to choose the cards that need to be turned over in order to determine whether the following rule is true or false: "If a card has a vowel on one side, then it has an even number on the other side." Most subjects choose ***E*** alone, or ***E*** and ***4***, while the correct answer is ***E*** and ***7***, because "Any odd number on the other side of ***E*** falsifies the rule in exactly the same way as would any vowel on the other side of ***7***" [Wason and Johnson-Laird, 1972]. According to First-Order Predicate Logic (FOPL), the "rule" to be tested is *(for-all x) (Vowel(x)* $\rightarrow$ *Even(x))*. This proposition is false if there is at least one card that satisfies *Vowel(x)* but not *Even(x)*, otherwise it is true. Therefore, to determine the truth value of the rule means to check all cards that may *falsify* the rule. What the subjects show is a tendency to find cards that *verify* the rule. In FOPL, cards verifying the rule make no contribution to its truth value. If the ***4*** card indeed has a vowel on the other side, it does not mean the rule is true. If the ***E*** card and ***7*** card do not make the rule false, the rule is true, and the ***4*** card does not need to be turned over.

This interpretation of the rule in FOPL is clearly different from what the subjects do in the task. This difference is usually judged to be an error in the reasoning of the subjects, not an error in logic, because "logic" in general, and FOPL in particular, is *normative model* of reasoning, which is justified according to certain universally accepted principles of rationality. Therefore, "a violation of logic" is just like a definition of "error". For this concrete situation, there are other arguments supporting the "human error" judgment. After the results are explained to the subjects, most of them agree that the correct answer is ***E*** and ***7***, and the results get better in variations of this task, such as when the subjects face a "reverse task" where the correct answer is given at the beginning, and the subjects are asked to explain why they are correct [Wason and Johnson-Laird, 1972]. A well-known variation is to change the artificial and abstract setting of the original task to realistic-content tasks, such as to ask the subjects to think themselves as police officers enforcing the following regulation: "If a person is drinking beer, then the person must be over 19 years of age". This situation has a one-to-one correspondence to the original task, only that here the four cards become four people: (1) someone drinking beer, (2) someone drinking Coke, (3) someone 21 years old, and (4) someone 16 years old. "To turn over a card" becomes "to check a person's age" (for the first two) or "to check what a person is drinking" (for the last two). When facing this task, subjects do "follow logic", that is, they check the age of person 1, and what person 4 is drinking [Griggs and Cox, 1982]. This result is often interpreted as "though people have difficulty in following logic in abstract reasoning, they can do so in concrete situations".

If human reasoning does not follow logic, then what does it follow? There are theories like mental logic [Braine and O'Brien, 1998], mental model [Johnson-Laird, 1983], and so on. Unlike logic, these theories are descriptive, not normative, and they are either completely informal, or semi-formal, that is, not as formal as a mathematical logic like FOPL. As Wason and Johnson-Laird said [1972]: "Only gradually did we realize first that there is no existing formal calculus which correctly modeled our subjects' inference, and second no purely formal calculus would succeed".

NARS, A NEW LOGIC

NARS (Non-Axiomatic Reasoning System) is an AI system that adapts to its environment and works with insufficient knowledge and resource [Wang, 1995]. In this paper, we only briefly introduce the aspects of the system that are most directly related to Wason's selection task. Interested reader should go to the web page of the author for related publications and an on-line demonstration of NARS (which will be presented in the Poster Session of ICCS2001).

In NARS, the basic form of knowledge is an *Inheritance relation*, $S \subset P$, from a *subject term* S to a *predicate term* P. Intuitively, it means that "S is a specialization of P, and P is a generalization of S". For example, *bird* $\subset$ *animal* roughly corresponds to "Bird is a kind of animal" in English. For a term T, its *extension* and *intension* are defined as sets of terms $E_T = \{x \mid x \subset T\}$ and $I_T = \{x \mid T \subset x\}$, respectively. Intuitively, they include all known specialization (instances) and generalizations (properties) of T, respectively. Since Inheritance relation is reflexive and transitive by definition, it can be proven that $(S \subset P) \Leftrightarrow ((E_S \subseteq E_P) \wedge (I_P \subseteq I_S))$, i.e., there is an Inheritance relation from S to P if and only if S inherits the intension of P, and P inherits the extension of S.

Given the assumption of insufficient knowledge, $S \subset P$ is usually uncertain, and the uncertainty can be measured according to available evidence. For a statement $S \subset P$ and a term M, if M is in the extensions of both S and P, it is positive evidence; if it is in the extensions of S but not the extension of P, it is negative evidence. Symmetrically, if M is in the intensions of both P and S, it is positive evidence; if it is in the intension of P but not the intension of S, it is negative evidence. Therefore, the amount of positive evidence is $w^+ = |E_S \cap E_P| + |I_P \cap I_S|$, the amount of negative evidence is $w^- = |E_S - E_P| + |I_P - I_S|$, and the amount of all evidence is $w = w^+ + w^- = |E_S| + |I_P|$.

The truth value of $S \subset P$ consists of a pair of numbers in [0, 1], $<f, c>$, where f is the *frequency* of positive evidence, $f = w^+/w$, and c is the *confidence*, a measurement on the amount of available evidence, $c = w/(w+1)$. Various inference rules are defined in NARS, including deduction, abduction, induction, revision, analogy, and so on. Each rule has a truth-value function to calculate the truth-value of the conclusion according to those of the premises. For our current purpose, the directly relevant rules are induction and revision (which are discussed in detail in [Wang, 1995]).

The induction rule derives a general conclusion from a single piece of evidence:

Premises: $M \subset P <f_1, c_1>; \quad M \subset S <f_2, c_2>$

Conclusion: $S \subset P \; <f_1, \; f_2c_1c_2 / (f_2c_1c_2+1)>$

The revision rule merges evidence from separate sources together:

Premises: $S \subset P <f_1, c_1>; \quad S \subset P <f_2, c_2>$

Conclusion: $S \subset P \; <f, c>$

where
$$f = [f_1c_1(1-c_2)+f_2c_2(1-c_1)] \, / \, [c_1(1-c_2)+c_2(1-c_1)]$$
$$c = [c_1(1-c_2)+c_2(1-c_1)] \, / \, [c_1(1-c_2)+c_2(1-c_1)+(1-c_1)(1-c_2)]$$

Let us use *EvenCard* to represent cards with an even number on its number side, and *VowelCard* for cards with a vowel on its letter side. Therefore, the “rule” to be evaluated in Wason's selection task can be written as an Inheritance statement *VowelCard* $\subset$ *EvenCard* in NARS. Given the above definitions, if a card is both a *VowelCard* and an *EvenCard*, it is a piece of positive evidence for the statement (according to the induction rule). If a card is a *VowelCard* but not an *EvenCard*, it is a piece of negative evidence for the statement (also according to the induction rule). In both cases, after finding such a card, the truth value of the rule is adjusted (according to the revision rule). If a card is not a *VowelCard*, it is not evidence for the statement, so the truth value remains the same, and whether it is an *EvenCard* does not matter.

Let’s see the four cards displayed in the task. (1) The ***E*** card is a *VowelCard*, so is evidence for the rule. If it has an even number on the other side, it is positive evidence, otherwise it is negative evidence. (2) The ***K*** card is not a *VowelCard*, therefore is irrelevant. (3) The ***4*** card is an *EvenCard*. If it has a vowel on the other side, it is positive evidence, otherwise it is irrelevant. (4) The ***7*** card is not an *EvenCard*. If it has a vowel on the other side, it is negative evidence, otherwise it is irrelevant.

Therefore, according to NARS, if the subject has considered all the possibilities and there is no other factor, then the correct decision is to turn over the ***E***, ***4***, and ***7*** card, but not the ***K*** card. However, NARS assumes that the subject works with insufficient knowledge and resources, therefore may fail to consider certain possibility. Also, there may be other factors that influence the decision, therefore we cannot expect such a system to make perfect decisions. Under the assumption that these factors equally affect all the possibilities, we can still expect different probabilities for the cards to be selected. The ***E*** card should be selected most often, because it corresponds to two possible ways to provide evidence. On the contrary, the ***K*** card should be selected least often, because it is not evidence, and its selection is only caused by the random factors. Between the ***4*** card and the ***7*** card, though they are logically symmetric, there are at least two factors that in favor of the former: one is the extra step of negation needed by the ***7*** card to be recognized as a piece of negative evidence, and the other is the priming effect on even numbers caused by the mentioning of "even number" in the description of the task. Adding them together, the ***4*** card should be selected more often than the ***7*** card. In summary, the experiment results are consistent with NARS.

DISCUSSION AND CONCLUSION

There are many differences between NARS and traditional formal logic, such as FOPL. For the current issue, the most relevant one is the relationship between "evidence" and "truth value" for the statement under testing. As discussed previously, in FOPL, "true" means "no negative evidence", and "false" means "has negative evidence", while whether there is positive evidence has no effect on the truth value of a statement. In NARS, both positive evidence and negative evidence contribute to truth value of a statement. Because negative evidence usually needs additional processing, it may be less accessible than positive evidence when the system has limited resources. These two types of logic are based on different foundations, and useful for different purposes. FOPL is *axiomatic* in the sense that it assumes sufficient knowledge and resources, so is more suitable for inference in a closed domain with conventional knowledge, such as in mathematics. On the contrary, NARS is *non-axiomatic* in the sense that it assumes insufficient knowledge and resources, so is more suitable for inference in an open world, where new evidence can be anything.

In Wason's selection task, the expected results are the ones of an axiomatic logic, while the actual results are consistent with a non-axiomatic logic. Therefore, the "mistake" here is mainly the misunderstanding between the psychologists who run the tests and the subjects who take the tests. In this artificially structured experiment, it is valid for the psychologists to assume sufficient knowledge and resources, therefore to expect the application of an axiomatic logic. Their mistake, however, is the failure to see the result as coming from another type of logic. On the side of subjects, since non-axiomatic logics are used more often in everyday life, they fail to understand the experiment setting as a testing of their capacity of using an axiomatic logic.

This explains why many subjects admit their mistake afterwards, and do better in the reverse task --- as soon as they realized that the expected logic is not their default one, they have less problem to follow it. It is similar in the concrete version of the task, where the instruction makes it clear that the task is checking for possible negative evidence of a given statement. Also, helped by the familiarity of the realistic situation, possible negative evidence becomes more recognizable. These factors make the subjects do better in this type of task.

Different from the other theories on reasoning, such as mental logic and mental model, NARS is a formal normative model of reasoning, justified according to certain theoretical assumptions, not a descriptive one, justified according to psychological data. Even so, since its assumptions (the insufficiency of knowledge and resources) are closer (compared to those of axiomatic logics) to the situation in which the human mind has evolved, it is not a surprise that it matches better with human reasoning.

Given the length restriction, this paper cannot provide a comprehensive coverage of either the selection task or NARS. The major claim made here is that in the previous study, the possibility of another type of logic has not obtained sufficient consideration. This paper is similar to [Oaksford and Chater, 1994] in attributing the problem revealed by the selection task not to human reasoning, but to traditional logic. The difference between these two is in the normative theory proposed for reasoning. For a comparison of NARS and probability theory, see [Wang, 1996].

In general, when a systematic difference is observed between human behaviors and a normative theory (such as FOPL and probability theory), people tend to jump to the "human error" conclusion, without careful studying the assumptions behind the normative theory, not to mention thinking about alternative normative theories whose assumptions are satisfied better by the given situation. This is a more serious human error than the so-called ones in previous study.

REFERENCES

M. Braine and D. O'Brien. (1998). *Mental Logic*. Lawrence Erlbaum Associates, Mahwah, NJ.

R. Griggs and J. Cox. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, **73**, 407-420.

P. Johnson-Laird. (1983). *Mental Models*. Harvard University Press, Cambridge, MA.

M. Oaksford and N. Chater. (1994). A Rational Analysis of the Selection Task as Optimal Data Selection. *Psychological Review*, **101**, 608-631.

P. Wang. (1995). *Non-Axiomatic Reasoning System: Exploring the Essence of Intelligence*. PhD Dissertation, Indiana University.

P. Wang. (1996). Heuristics and normative models of judgment under uncertainty. *International Journal of Approximate Reasoning*, **14**, 221-235.

P. Wason and P. Johnson-Laird. (1972). *Psychology of Reasoning: Structure and Content*. Harvard University Press.