When participants are not misled they are not so bad after all: A pragmatic analysis of a rule discovery task

Jean-Baptiste Van der Henst (jvanderhenst@hotmail.com)
Laboratory of Experimental Psychology, K.U. Leuven, Tiensestraat 102
Leuven, 3000 Belgium

Sandrine Rossi (rossi@scvie.unicaen.fr)
Laboratoire de Psychologie Cognitive et Pathologique, Université de Caen, Esplanade la Paix
Caen, 14032 France

Walter Schroyens (Walter.Schroyens@psy.kuleuven.ac.be)
Laboratory of Experimental Psychology, K.U. Leuven, Tiensestraat 102
Leuven, 3000 Belgium

Abstract

In this paper we present a pragmatic analysis of a widely used task in field of hypothesis testing: the 2-4-6 problem (Wason, 1960). In this task participants have to discover the rule "three increasing numbers" by proposing triples and are given the "2-4-6" as an example of triples compatible with the rule. We argue that most people fail because the givens of the task are conversationally misleading: first because the 2-4-6 is communicated and is thus presumed to be relevant (Sperber & Wilson, 1995) and second because the rule to be discovered is too trivial in the context of the task. In a first experiment we showed that providing the triple without communicating it improved performance in the task. In a second experiment we contextually increased the relevance of such a rule and observed that people were thus more inclined to discover it.

Introduction

Imagine that you have to discover a rule that generates triples of numbers. Some triples are consistent with the rule and some are not. Now, somebody who knows the rule – a trustworthy person like an experimental psychologist – is telling you that '2-4-6' is a triple that is consistent with the rule. Will you consider this as helpful information or not? Surely, you will and you will also probably think that the experimental psychologist would expect you to regard this triple as helpful in order to succeed in the task. Hence, you will consider that the salient properties conveyed by '2-4-6' (like "evenness" or "increase by 2") must be taken into account in order to discover the rule. However, considering that these properties are important is in fact deceptive since the rule to be discovered does not relate to them: the rule is simply *"three increasing numbers"*. Focusing on 2-4-6's most salient properties is thus not the good way to solve such a task!

This task is the well-known '2-4-6' problem designed by Peter Wason more than forty years ago (Wason, 1960) in order to investigate hypothesis testing ability (see Gorman, 1995 and Poletiek, 2001 for reviews). It has become the most commonly used task by researchers in the field of hypothesis testing. In its standard version, it consists in proposing sequences of triples to discover a rule the experimenter has in mind. For each triple, the experimenter indicates whether or not it is consistent with the rule. Participants have to test triples until they are sure of having discovered the rule. As for the other famous Wason's task, namely the selection task (Wason, 1968), one stimulating aspect of the '2-4-6' problem is that few people succeed in it despite of its apparent simplicity. In the initial study (Wason, 1960), only 21% of participants succeeded in discovering the rule in their first announcement. Typically, the rules proposed by participants inherit the salient properties of '2-4-6' and are more specific than the rule to be discovered. For instance, they propose rules such as *"three even numbers", "numbers increasing by 2", "even numbers increasing by 2"*.

The failure in the 2-4-6 task has often been viewed as a sign of irrationality. Wason argued that participants exhibited to a "confirmation" bias, and Evans (1983; 1989) argued that people exhibited to a "positivity" bias. The protocols indeed reveal that people tend to propose instances of triples compatible with their hypothesis whereas the most efficient strategy consists of proposing instances inconsistent with the held hypothesis. It is commonly accepted that people overly rely on a positive testing strategy and focus on too narrow hypotheses (Poletiek, 2001). What is the reason for this? Researchers assume that positive testing is simply a natural way of thinking typical of human beings (Evans, 1989; Klayman, 1995) and that the consideration of restrictive hypotheses is due to the salient properties of 2-4-6. In contrast, we think that the

incidence of positive testing and the size of "restrictiveness" bias (Poletiek, 2001) have been overestimated. We believe that one of the most important analysis required for understanding the psychological mechanisms underlying the task has been systematically overlooked, namely the *pragmatic analysis*. In the present paper, we argue that a key cause of the failure in the 2-4-6 problem derives from *communication.* In particular, we claim that the task is difficult precisely because the triple 2-4-6 is obtained by communication.

The pragmatics of the 2-4-6 problem

Nobody can contest that conversation plays an important role in the task. First, like many reasoning tasks, the 2-4-6 problem sets up a situation of communication. The experimenter communicates the givens of a problem to a participant and the participant has to communicate the experimenter a conclusion in order to provide him with some information about his or her inferential skills. The participant tries to determine the experimenter's communicative intention and has some expectations about what he or she is interested in. She or he may thus tailor her/his answer according to these expectations. Second, and more importantly, communication is noticeably misleading in the 2-4-6 problem. In Gricean terms, one can view the experimenter as being uncooperative (Grice, 1975) since the triple he/she intentionally choose is overly specific and does not illustrate the level of generality of the rule. Communicating the triple '2-4-6' to illustrate a typical example of the rule is thus a violation of the second maxim of quantity which stipulates that the speaker should not make his/her contribution more informative than required (Grice, 1975). Of course choosing a triple whose most salient properties are consistent with much more specific rules than the one to be discovered was done in purpose. Wason and other subsequent researchers wanted to see if people were able to come up with general hypotheses by attempting to falsify specific hypotheses drawn from the triple 2-4-6. However, what has been neglected is the fact that the consideration of specific hypotheses is made on the basis of a triple that is communicated. Giving the participant a specific triple has not been seen as a violation of a rule of communication but rather as a way to suggest specific hypotheses. Consequently, researchers have not assessed the impact of misleading communication on weak performance in the task. To which extent does the fact that 2-4-6 is communicated contribute to the restrictiveness bias? In this paper, we aim at investigating this issue.

Our pragmatic analysis of the 2-4-6 relies on relevance theory (Sperber and Wilson, 1995). The concept of relevance is characterized by cognitive effects and cognitive effort, and the degree of relevance relies on these two factors: on one hand, the greater the cognitive effects resulting from processing an information, the more relevant the information; on the other hand, the greater cognitive effort required to achieve these effects and process that information the lesser its relevance. Sperber and Wilson (1995) argue that human communication is governed by a *communicative principle of relevance.* According to this principle, each utterance conveys a presumption of its own relevance. This makes an important difference between information received from a communicator and information not obtained by communication. A communicated information raises expectations of relevance. The communicator manifestly intends the information to be relevant enough to deserve the consideration by the addressee. Presuming that the information is relevant implies first that the effects will be sufficient to offset the effort required to process the communicated information. Second, it implies that the effort required is presumed to be minimal to reach the level of expected effects given the communicator's ability. The presumption of relevance sets up a comprehension strategy, which consists in following a least-effort path: considering cognitive effects in order of accessibility and stopping the processing effort when the level of expected relevance is met.

Let's now turn to the task itself. When the experimenter gives the participant '2-4-6' as an example consistent with the rule to be discovered, this triple is accompanied by a presumption of relevance. In other words, the addressee should presume that this triple is relevant to discover the rule. What type of cognitive effects may the addressee expect to draw from processing the triple? She/He will expect that the triple will look as having been generated by a rule and thus will search for properties common to the three numbers or for properties about the way these numbers are ordered. There are many properties that can be attributed to the 2-4-6 triple, but some of these properties are much more salient (i.e. they immediately come to mind with minimal processing) than others. Given the presumption of relevance, the properties to be considered in order to discover the rule are those that are easily accessible from processing 2-4-6 (for instance "evenness" and "increase by 2").

This does not necessarily mean that the rule should correspond exactly to one or several of the most salient properties but this means that these properties indicate the directions to investigate in order to discover the rule. Actually, the rule may still be quite hard to discover and integrate many other characteristics than the one conveyed by the triple, but what is saliently given in the triple is relevant to discover it. Hence, because the triple is presumed to be relevant, the most accessible properties it conveys cannot be considered as inappropriate clues to discover the rule. When the

experimenter communicates the triple '2-4-6' while he wants the participant to discover the rule "three increasing numbers", he violates the participants' expectations of relevance.

Another cause that may contribute to mislead participants is the rule to be discovered itself. When a person is taking part in a reasoning experiment she/he should normally expect to provide the experimenter with some information about her/his inferential skills. Trivial tasks and trivial answers are not well suited for this: they require little cognitive competency and should not interest the experimenter. In order to provide an answer relevant to the experimenter, the participant will probably expect the task to be of a certain level of difficulty. Faced with a task consisting of discovering a rule about numbers, participants should thus think that this rule is not a trivial one. However, "increasing numbers" is actually one of the most obvious rules applied to order numbers and in daily life we very often encounter set of numbers sorted according to it. Consequently, participants may be reluctant to think of discovering such an undemanding and widespread rule and may try to achieve relevance by seeking for non-trivial rules. As he/she is misleading in communicating the triple, the experimenter is also misleading about his own expectations since he actually expects the participants to find out one of the less relevant (i.e. the simplest) rule to discover in the context of the task.

In our experiments we aimed at showing that misleading participants in their expectations of relevance influence task performance. We used a less deceptive way to convey the givens of the problem while still keeping the same provided example (i.e. 2-4-6) and the same target rule (i.e. increasing numbers). In the first experiment, the presumption of relevance, which accompanies the triple 2-4-6 in the standard version of the task, was removed: 2-4-6 was *given* but *not communicated*. In the second experiment, we increased the relevance of the target rule "three increasing numbers": We did not make the search for the rule less trivial but we framed the task within a context in which discovering such a rule became much more relevant.

Experiment 1

In this experiment, there was no presumption of relevance accompanying the triple 2-4-6 in one of the two conditions, since the participant did not obtain it by communication. Before receiving the instructions of the rule discovery task, participants had to manipulate a "jackpot" generating triples of numbers at random. After several trials, the experimenter gave the participant the instructions about the rule discovery task. He/She then asked the participant to trigger the jackpot for a last time and told him whether or not the obtained triple was compatible with the rule. However, the jackpot was biased in such a way that the sequence 2-4-6 came out on this trial. The participant obviously did not know that we rigged the jackpot on this trial and could expect the triple to be consistent or inconsistent with the rule. Hence, from the participant's perspective, the salient properties of the 2-4-6 just result from chance. Even if this triple suggests specific hypotheses, the participant cannot consider them as ones the experimenter necessarily wanted him/her to think about. This because the triple has not been chosen intentionally, in contrast with the standard version of the task. Participants should rely much less on the salient properties of the 2-4-6 and should perform better than in the standard version. We predict that subjects would focus less on the specific properties conveyed by the triple. Consequently, they should test a greater variety of triples: we should thus observe a greater rate of triples increasing in an irregular way (i.e. triples whose numbers do not increase with the same interval, see also Vallée-Tourangeau, Austin & Rankin, 1995) or counterexamples of the rule to be discovered (i.e. triples which are not increasing).

Participants

Fifty-eight undergraduate psychology students from the University of Caen (France) volunteered to participate in this experiment for a course requirement. They were tested individually.

Procedure and materials

In the control condition (N=29) participants received the task with the standard instructions. They were required to discover a rule the experimenter had in mind by proposing sequences of three numbers and were given '2-4-6' as an example of triples compatible with the rule. To make sure that the participant well understood the instructions, which were printed on an instruction sheet, the experimenter re-explained them and asked the subject if he/she had any questions. Participants kept a written record of the triples they proposed, their hypothesis about the target rule, as well as the experimenter feedback.

In the "jackpot" condition (N=29), participants were faced with a computer screen resembling a jackpot machine. Participants were informed that it randomly generates sequences of three numbers. The experimenter asked the participants to trigger the jackpot by pressing the key "ENTER". After five trials, the experimenter stopped the "jackpot" session. At this point, the participant did not know yet the purpose of the task and did not know what the use of the jackpot was for. After the jackpot session, the participant was given the rule discovery task as in the standard version. However, instead of receiving an example communicated by the experimenter, he/she had to

trigger the jackpot for a last trial. The experimenter would thus tell him/her whether or not the triple supposed to be randomly generated was consistent with the rule. For each participant, the sequence 2-4-6 came out at this trial and the experimenter told the subject that it was consistent with the rule to be discovered. The participant subsequently had to generate numbers by herself/himself, like in the standard version.

Results

Performance As predicted participants performed better in the jackpot condition than in the control condition: Only 24% of participants gave the correct rule on their first announcement in the control condition (21% in Wason' study) while 55% did so in the jackpot condition (χ^2 (1) = 5.84, p <.02). Moreover, the mean number of rules announced to reach the correct solution was higher in the control condition than in the jackpot condition (2.38 vs. 1.59; *Mann-Whitney* $U_{29,24}$= 214, Z=2.39, p<.01) ; for the five participants of the control group who failed in the task, the mean number of proposed rules was 2.6).

Number and types of triples Participants tested more triples before proposing a rule in the jackpot condition than in the control condition. The mean number of proposed triples per rule by participants who succeeded in the task was higher in the jackpot condition than in the control condition (8.15 vs. 6.11; $U_{29,24}$=202, Z=-2.6, p<.009). The mean proportion of counter-examples (i.e. triples that received negative feedback) for successful subjects was lower in the control condition than in the jackpot condition (0.17 vs. 0.25; $U_{29,24}$=207.5, Z=-2.51, p<.01; this rate is equal to 0.06 for the 5 participants who failed in the control condition). Similarly, the mean proportion of irregular increasing triples was lower in the control condition than in the jackpot condition for these subjects (0.18 vs. 0.29; $U_{29,24}$=218, Z=-2.32, p<.02; this rate is equal to 0.02 for the failing participants). These results indicate that subjects in the jackpot conditions were more prompt to explore a greater variety of triples than subjects in the control condition who focused more on triples exhibiting the salient properties of the 2-4-6.

Discussion

Removing the presumption of relevance of '2-4-6' was helpful: twice as many subjects immediately discovered the correct rule when the triple 2-4-6 was provided without any presumption of relevance and 83% of participants in the jackpot condition succeeded in the task in no more than two announcements. Moreover, even though giving 2-4-6 in the jackpot condition may still suggest specific hypotheses related to the salient features of this triple, the fact of removing its presumption of relevance entails that these features do not necessarily have to be taken into account in order to succeed in the task. Consequently, participants in the jackpot condition were more inclined to consider alternative properties and thus tested a greater variety of triples than in the control condition.

Experiment 2

In contrast with the Wason selection task literature, in which the content question has led to a vast amount of research, no study has ever investigated content effects with the 2-4-6 problem. That is in the earlier experiments numbers in the triples proposed by participants or given by the experimenter never referred to real quantities. However, when we use numbers in daily life, we most of the time refer to concrete quantities like books, people, dollars, reasoning errors and so on. We conjectured that framing the task with a real content situation might influence performance. In particular, we thought that using an appropriate content could enhance the relevance of searching for the rule "three increasing numbers". In many real life situations, following a rule of increase is actually highly relevant. A paradigmatic example is economic activity. An economic agent always aims at following such a rule: he or she wants his/her turnover, sailings, productivity, profits, or market scope to increase over the time. Hence, we framed the 2-4-6 task in the context of economic activity. In such a context, the task was to discover not a rule an experimenter − who studies human reasoning and hypothesis testing − had in mind but a rule about car sailings that a garage owner imposes to his new employee. From the participant's perspective, what is relevant for an experimental psychologist who studies cognitive skills is likely to be different from what is relevant for a garage owner. Indeed, searching for the most common way to order numbers (i.e. rule of increase) may be seen as trivial in one case, whereas it becomes highly relevant in the other. Hence, we predict that people will discover the rule "three increasing numbers" more often in the "economic" condition than in the control condition.

Participants

One hundred and twenty undergraduate psychology students from the University of Leuven (Belgium) volunteered to participate in this experiment for a course requirement.

Procedure and materials

The procedure used in this experiment substantially differed from the one we used in Experiment 1. Indeed, participants did not interact with the experimenter. The instructions were given via a computer and participants had to enter the triples they wanted to propose in the

computer. The feedback about the triple proceeded in the following way: when the proposed triple was consistent with the rule it appeared in green and when it was inconsistent it appeared in red. The consistent/inconsistent triples remained on screen where they were presented one beneath the other in different columns and the order they were presented. After each trial participants had the press the [1] key when they wanted to test another triple, and the [2] key when he/she was sure of having discovered the rule, after which he/she had to type in the rule. After one rule announcement, the experiment ended. This contrasts with Experiment 1 in which participants could propose rules until they discovered the target rule. Participants were tested in groups of ten to twenty people each on their individual computer. In the control condition (N=62), participants had to discover a rule implemented by the experimenter in the computer. In the "economic" condition (N=58) participants had to discover a rule a garage owner imposes to his new employee.

Results and discussion

In line with our prediction, more participants discovered the rule in the economic condition than in the control condition (3.2% vs. 29.3%; $\chi^2(1)= 15.3$, $p=.0001$). Moreover participants proposed a greater number of triples in the economic condition than in the control one (2.45 vs. 3.73; *Mann-Whitney* $U_{62,58}= 1121.0, Z = -3.645,$, $p <.05$). This indicates that people abstract more beyond the hypotheses suggested by the 2-4-6 triple and thus tested a greater variety of triples. In particular, the mean proportion of irregular increasing triples was higher in the economic condition than in the control one (.090 vs. .125, *Mann-Whitney* $U_{62,58}= 1553.5, Z = -1.689, p < .05$ one-tailed). The mean rate of counter-examples was also higher in the economic condition than in the control one but this difference was not significant (.10 vs. .12). Hence, the results reveal that increasing the relevance for searching the rule "three increasing numbers" improved performance and thus show that people adapt their cognitive skills in order to maximize relevance (Sperber & Wilson, 1995): on one hand, it is indeed not really relevant for a participant to search for a trivial rule like "increasing numbers" in order to exhibit to a cognitive psychologist her/his own cognitive skills, especially if the experimenter provides an example which does not suggest such a rule; on the other hand, it becomes much more relevant to search for a rule of plain increasing in the context of economic activity.

General discussion

In this study we provided the first extensive conversational analysis of the 2-4-6 problem. We argued that communication was highly misleading in this task and that this explains why people focus on overly narrow hypotheses. We claimed that being misled is not a clue of irrationality. We showed that when participants were not misled, they were not so bad after all. In the standard task, people are misled because they rationally consider that the givens of the problem are relevant to solve the task.

Our experiments aimed at providing less deceptive tasks and show that this increased performance. In Experiment 1, we designed a task in such a way that there was no presumption of relevance accompanying the triple 2-4-6. This implied that the salient properties of the 2-4-6 did not have necessarily to be considered by the participant in order to succeed in the task. The results showed that participants performed better when the salient characteristics of the 2-4-6-triple resulted from a random procedure (a jackpot) than when a presumption of relevance accompanied such a triple (as in communication). In, Experiment 2 we manipulated the content of the task: triples did not refer to abstract numbers as it has always been the case in previous studies but to numbers of cars sold within a three months period. We framed the task within the context of economic activity in order to increase the relevance of searching for the rule "three increasing numbers". Our study shows that participants tailor the search of their hypotheses according to what they expect to be relevant in the task. When expectations of relevance coincide with the correct rule, they are more prone to discover it. This is line with the pragmatic analysis of the Wason selection task made by Sperber, Cara and Girotto (1995). According to them, performance in this task is determined by expectations of relevance. Subjects fail in this task because intuitions of relevance do not coincide with the logical answer. Increasing the rate success consist in constructing a context in which intuitions of relevance will match with the logical answer.

A pragmatic analysis of the conversational structure of a cognitive task may be highly helpful in assessing the quality of participants' skills. There is now an increasing body of research in the domain of high-level cognitive processes revealing that lack of rationality is mistakenly attributed on the basis of misleading tasks (for reviews see, Politzer, 1986; Hilton, 1995; Politzer & Macchi, 2000). These studies as well as ours should alert psychologists that weak performance in a task might be overestimated in the absence of a pragmatic analysis.

Acknowledgements

We are very grateful to Professor D. Trotoux and his students of the University Institute of Technology of Caen for the design of the "jackpot" we used in Experiment 1.

References

Evans, J. St. B. T. (1983). Selective processes in reasoning In J. St B. T. (Ed.), *Thinking and Reasoning: Psychological Approaches*. London: Routledge & Kegan Paul.

Evans, J.St.B.T. (1989). *Bias in human reasoning : Causes and consequences*. Hove, UK : Lawrence Erlbaum Associates.

Grice, H.P. (1975). Logic and conversation. In P. Cole, & J.L. Morgan (Eds), *Studies in syntax, Vol 3: Speech acts*. New York: Academic Press.

Gorman, M. E. (1995). Hypothesis testing In S. E. Newstead & J. St. B. T. Evans (Eds.), *Perspectives on Thinking and Reasoning. Essays in honour of Peter Wason*. Hove, UK: Lawrence Erlbaum Associates.

Hilton D.J. (1995) The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin*, *118*, 248-271.

Klayman J., Ha Y.W. (1987) Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review*, *94*, 211-228.

Klayman, J. (1995). Varieties of Confirmation Bias In J. R. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Decision Making from the Perspective of Cognitive Psychology* . New York: Academic Press.

Poletiek, F. (2001). *Hypothesis-testing behaviour*. Psychology Press/Taylor and Francis, Philadelphia.

Politzer G. (1986) Laws of language use and formal logic. *Journal of Psycholinguistic Research*, *15*, 47-92.

Politzer, G. & Macchi, (2000) L. Reasoning and pragmatics. *Mind and Society*, *1*, 73-93.

Sperber D., Cara F., Girotto V. (1995) Relevance theory explains the selection task. *Cognition*, *57*, 31-95.

Sperber, D. & Wilson, D. (1995). *Relevance: Communication and Cognition,* Oxford, Blackwell. 2nde Edition.

Vallée-Tourangeau, F., Austin, N. G., & Rankin, S. (1995). Inducing a Rule in Wason's 2-4-6 Task: A Test of the Information-Quantity and Goal-Complementarity Hypotheses. *Quarterly Journal of Experimental Psychology, 48,* 895-914.

Wason, P.C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12*, 129-140.

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