Memory for a short sequence of assignment statements (part 2 of 2)

Derek M. Jones
derek@knosof.co.uk

1 Introduction

This is the second of a two part article describing an experiment carried out during the 2004 ACCU conference. The previous part was published in the previous issue of C Vu. This second part discusses how the `if` statement part of the problem affected subject performance.

The `if` statement problem can be viewed as either a time filler for the assignment remember/recall problem, or as the main subject of the experiment (with the assignment problem acting as a smoke screen to make it more difficult for subjects to notice any patterns in the `if` problems). The reason for this second possibility is that studies have found patterns in the errors made by subjects when performing various kinds of deduction tasks.

Given that some kind of filler task had to be performed, your author decided to take opportunity to try and replicate some of the errors patterns seen in some studies of deduction.

As Table 1 shows, relational operators commonly occur in `if` statements.

Table 1: Occurrence of equality, relational, and logical operators in the conditional expression of an `if` statement (as a percentage of all such controlling expressions and as a percentage of the respective operator). Based on the visible form of over 3 million lines of C source. The percentage of controlling expressions may sum to more than 100% because more than one of the operators occurs in the same expression.

Operator	% Controlling Expression	% Occurrence of Operator	Operator	% Controlling Expression	% Occurrence of Operator
==	31.7	88.6	>=	3.5	76.8
!=	14.1	79.7	no relational/equality	47.5	—
<	6.9	45.6	||	9.6	85.9
<=	1.9	68.6	&&	14.5	82.3
>	3.5	84.9	no logical operators	84.2	—

2 Linear syllogisms

The psychology of deduction uses the terms *linear syllogisms* or *linear reasoning* to describe deduction between statements involving relational operators. The term usually used to describe a (sub)expression containing a relational operator, in programming language specifications, is *relational expression*.

Linear syllogisms are part of mathematical logic and the skills associated with being able to make deductions based on relational information are usually assumed to simply be a component of the general reasoning ability that people have. However, studies have found that a number of animals have the ability to adapt their behavior to given situations based on relational knowledge they have acquired. For instance, aggressive behavior may occur between two animals to determine which is dominant, relative to the other. Such behavior can lead to being injured in a fight and is best avoided if possible. The ability to make use of relative dominance information (e.g., obtained by a member of a social group watching the interaction between other members of the group) may remove the need for aggressive behavior during an encounter between two members of the same group who have not met face to face before (i.e., the member most likely to loose immediately behaves in a subservient fashion).

One study[11] allowed a social dominance hierarchy to become established in several independent groups of birds (Pinyon jays). Two birds from different groups were then placed in a cage and given time to establish their relative social dominance (a process that involves staring, looking away, chin-up and beg, etc). The interaction of the two birds was witnessed by a bird belonging to one of the two groups from which the two birds came (this bird could not participate in any social interaction with the birds it witnessed). The witness bird had previously encountered one of the bird in the interaction it witnessed, but had never seen the other before. The witness bird was then allowed to interact with the bird from the other group. Analysis of the social interaction that occurred between the two birds on their first encounter showed that in those cases where the witness bird had sufficient information to reliably deduce its relative social status, it more often behaved in a way consistent with that social position, than an experimental control that had not witnessed

any interaction.

The results from a related study using western scrub jays (a less social species, closely related to pinyon jays) showed less evidence for the ability to make use of relational information. Those animals that live together in social groups are likely to have various kinds of relational information available to them. The benefits of being able to make use of this information appears to have resulted in at least some social species developing the cognitive abilities needed to process and make use of this information.

2.1 Relational reasoning in humans

If some animal brains (that don't have what are considered higher level cognitive reasoning abilities) have developed a mechanism to combine relational information to create *new* information, it is possible that humans also possess a similar mechanism (this is not to say that they don't have any other cognitive systems that are capable of performing the same task). A possible consequence of having such a special purpose reasoning mechanism is that it may not handle all relational expressions in the same way (i.e., it is likely to be optimized for handling those situations that commonly occur in it's owners everyday life). Some of the studies of human linear reasoning have found that subjects are slower and make more errors when the operands in a sequence of relational expressions occur in certain orders.

One study[4] used a task that was based on what is known as *social reasoning* (using the relations *better* and *worse*). Subjects were shown two premises, involving three people, and a possible conclusion (e.g., *Is Mantle worse than Moskowitz?*). They had 10 seconds to answer "yes", "no", or "don't know". All four possible combinations of conclusions were used.

Table 2: Eight sets of premises describing the same relative ordering between A, B, and C (peoples names were used in the study) in different ways, followed by the percentage of subjects giving the correct answer. Adapted from De Soto, London, and Handel.[4]

	Premises	Percentage Correct Responses		Premises	Percentage Correct Responses
1	A is better than B B is better than C	60.5	5	A is better than B C is worse than B	61.8
2	B is better than C A is better than B	52.8	6	C is worse than B A is better than B	57.0
3	B is worse than A C is worse than B	50.0	7	B is worse than A B is better than C	41.5
4	C is worse than B B is worse than A	42.5	8	B is better than C B is worse than A	38.3

Based on the results (see Table 2) the researchers made two observations (which they called *paralogical principles*; cases 5 and 6 possess both, while cases 7 and 8 possess neither):

1. People learn orderings better in one direction than another. In this study people gave more correct answers when the direction was better-to-worse (case 1), than mixed direction (case 2, 3), and were least correct in the direction worse-to-better (case 4). This suggests that use of the word *better* should be preferred over *worse* (the British National Corpus[9] lists *better* as appearing 143 times per million words, while *worse* appears under 10 times per million words and is not listed in the top 124,000 most used words).

2. People end-anchor orderings. That is, they focus on the two extremes of the ordering. In this study people gave more correct answers when the premises stated an end term (better or worse) followed by the middle term, than a middle term followed by an end term.

A related experiment in the same study used the relations *to-the-left* and *to-the-right*, and *above* and *below*. The above/below results were very similar to those for better/worse. The left-right results showed that subjects performed better with a left-to-right ordering than a right-to-left ordering.

Since this original study additional factors have been discovered and a number of models have been proposed to explain the strategies used by people in solving linear reasoning problems, including:

- *The spatial model[4, 7]*, in which people integrate information from each premise into a spatial array representing all known relationships.

- *The linguistic model[31]*, in which people represent each premise using linguistic propositions (the individual premises are not integrated).

- *The algorithmic model[13]*, in which people apply some algorithm to the structure of the linguistic representation of the premises. For instance, given "Reg is taller than Jason; Keith is shorter than Jason" and the question "Who is the shortest?", a so called *elimination strategy* was used by some subjects in the study. (The answer for the first premise is Jason, which eliminates Reg; the answer to the second premise is Keith which eliminates Jason, so Keith is the answer).

- *The mixed model[15]*, in which the information in the premise is first decoded into a linguistic form and then encoded into a spatial form.

The strategy used to solve a given problem has been found to vary between people. A study by Sternberg and Weil[16] found a significant interaction between a subjects' aptitude (as measured by verbal and spatial ability tests) and the strategy they used to solve linear reasoning problems. However, a person having high spatial ability, for instance, does not necessarily use a spatial strategy. A study by Roberts, Gilmore, and Wood[14] asked subjects to solve what appeared to be a spatial problem (requiring the use of a very inefficient spatial strategy to solve). Subjects with high spatial ability used non-spatial strategies, while those with low spatial ability used a spatial strategy. The conclusion made was that those with high spatial ability were able to see that the spatial strategy was inefficient to select as alternative strategy, while those with less spatial ability were unable to perform this evaluation.

If the evaluation of relational expressions in source code is performed using a cognitive mechanism that has been optimized for certain kinds of operations, then it is possible that developers performance will be worse for some forms of expressions (e.g., the rate of making mistakes will be greater). The form of the **if** statements used in this study was designed to look for differences in subject performance that depended on the form of the relational expressions appearing in the control expressions.

3 Subject motivation

When reading source code developers are aware that some pf the information they see only needs to be remembered for a short period of time, while other information needs to be remembered over a longer period. For instance, when deducing the affect of calling a given function the names of identifiers declared locally within it only have significance within that function and there is unlikely to be any need to recall information about them in other contexts. Each of the problems seen by subjects in this study could be treated in the same way as an individual function definition (i.e., it is necessary to remember particular identifiers and the values they represent, once a problem has been answered there is no longer any need to remember this information).

Subjects can approach the demands of answering the problems this study presents them in a number of ways, including the following:

- seeing it as a challenge to accurately recall the assignment information (i.e., minimizing *would refer back* answers),

- recognizing that *would refer back* is always an option, but that it is more important to correctly answer the **if** statement question,

- making no conscious decision about how to approach the answering of problems,
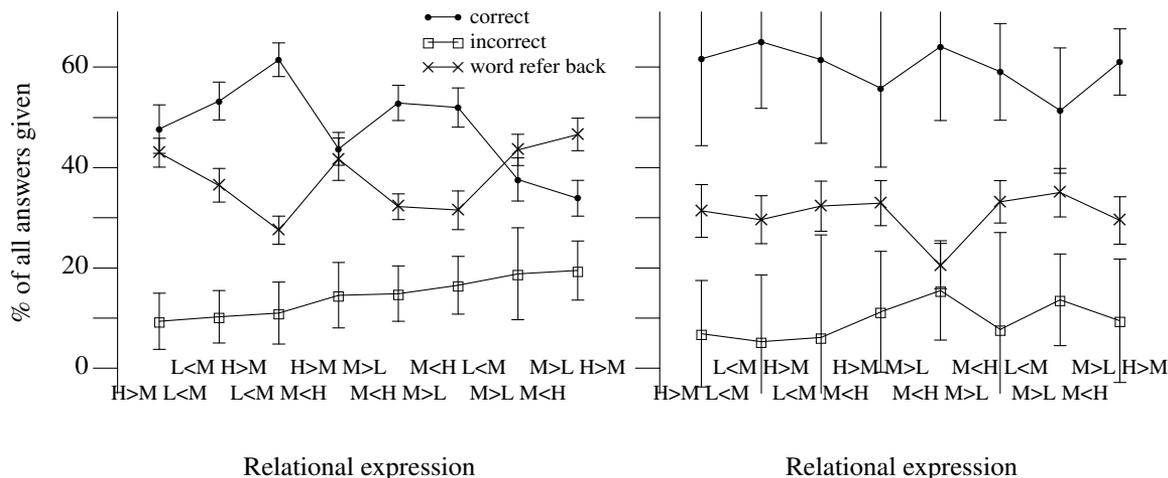
Figure 1: The percentage of *would refer back*, correct and incorrect answers for each kind of relational expression. The left graph is based on answers to the first eight problems, while the right graph is based on the answers from the ninth and subsequent problem answers. Variation in subject performance is denoted by the error bars, which encompass one standard deviation. The ordering of relational expressions along the x-axis is sorted on the percentage of incorrect answers to the assignment problem, for the first eight **if** statement problems. *H* denotes high, *M* denotes middle, and *L* denotes low. So "H > M M > L" denotes "high greater than middle and middle greater than low".

Experience shows that many developers are competitive and that accurately recalling the assignment information, after solving the **if** statement problem, would be seen as the ideal performance to aim for. The experimental format did not allow for easy debriefing of subject after they had answered the questions, and none was performed.

The only applicable instruction given to subjects was: "Read the variables and the values assigned to them as you might when carefully reading lines of code in a function definition."

4 Results

The raw results for each subject are available on the studies web page.[8]

4.1 if statement/assignment recall interaction

Answering the **if** statement portion of the problem requires time (information held in short term memory decays over time and unless it is regularly refreshed it will soon be lost) and use of short term memory resources. If subjects require more time or use more short term memory resources to answer some forms of relational expression problem, then performance in recalling assignment information is likely to be poorer after comprehending expressions having the *more complicated* form. The results (Figure 1) suggests that such a correlation may exist, at least for the first eight answers.

However, the difference in performance characteristics between the first eight answers and the ninth and subsequent problems may have been caused by subjects learning and making use of patterns in the assignment recall questions (which could reduce the need for short term memory resources). Alternatively some information occurred sufficiently often (e.g., the same identifier) that it was stored in a longer term memory subsystem, where it was not so susceptible to interference from the **if** statement problem.

4.2 if statement performance

This study differed from others on the topic of reasoning in a number of ways, including:

1. Researchers of human reasoning are usually attempting to understand the mechanisms underlying human cognition. For this reason they use subjects who have little or no experience in using formal
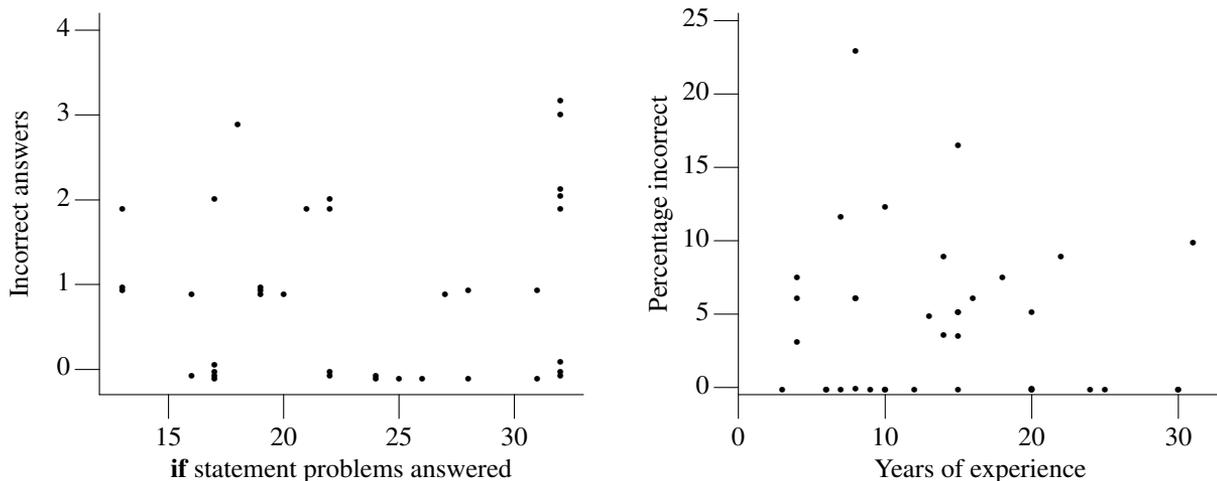
Figure 2: The left graph plots the number of problems answered by each subject against the number of incorrect answers they gave. The bullets are offset from the y-axis to try to show those cases where more than one subject had the same problems answered/incorrect answers pair. The right graph plots the number of years of subject experience against the percentage of incorrect answers they gave.

mathematical logic. This study was interested in the performance of subjects in evaluating particular kinds of logical expressions and subjects were chosen because they had significant amounts of experience in evaluating the kinds of logical expressions that occur in source code.

2. The problems used in studies by researchers investigating the mechanisms of human cognition are usually expressed in forms that occur in everyday life, i.e., they are natural language descriptions of everyday situations (e.g., "If Jim deposits 50p, he gets a canned drink."). One of the complications caused by expressing problems in this form is that the words and phrases used are often open to multiple interpretations. It is also possible that subjects will base their answer on expectations they have about how the *real world* operates.[5]

3. In this study no limits were placed on subjects (De Soto et al.[4] required that an answer be given within 10 seconds), the mode of presentation mimiced that encountered in program comprehension (in the Huttenlocher[7] study subjects heard a tape recoding of the problem).

A total of 844 **if** statement problems were answered. There were 40 (4.7% of all answers) incorrect answers, an average number of incorrect answers per subject of one. However, the incorrect answers were not evenly distributed across subjects. The number of incorrect answers did not appear to depend on the number of problems answered (Figure 2). While performance on reasoning tasks has been found to decrease with age,[6] years of experience (which is likely to be highly correlated with age) does not appear to have been a factor affecting the number of incorrect answers given to **if** statement problems.

Two the reasons why subject performance could differ across different forms of relational expressions are:

1. Subjects may have a cognitive *relational deduction* mechanism (this may be actual hardware, i.e., a cluster of brain cells, or software, i.e., a neural network whose weights have bee tuned through experience) that is optimized for handling problems (i.e., those that commonly occur in everyday life) that are expressed in a particular form.

2. The amount of cognitive resources required to solve a relational expression may depend on the form in which the expression is presented (this difference might simply be a consequence of how the human cognitive subsystem handles relational reasoning).

The paper and pencil format of the experiment meant that it was not feasible to obtain information on the amount of time taken to answer each problem.

Although subjects were told: "Treat the paper as if it were a screen, i.e., it cannot be written on.", there was nothing to prevent them using any paper that they happened to have on them as a temporary work area. Several subjects did write notes on the paper next to `if` statement problems (in one case for all the answered problems) and the answers to these problems were not counted. Except for the one case the number of such answers was very small (in the one case the subject was not included in the subject count).

Table 3: Number of correct and incorrect responses for the first eight and ninth and subsequent answers (parenthesized value is percentage of incorrect responses). *H* denotes high, *M* denotes middle, and *L* denotes low. So "H > M M > L" denotes "high greater than middle and middle greater than low".

relational form	correct (first 8)	correct (9th and subsequent)	incorrect (first 8)	incorrect (9th and subsequent)	incorrect (total)
H > M M > L	34	80	0 (0.0%)	2 (2.5%)	2 (1.8%)
L < M H > M	38	66	0 (0.0%)	3 (4.5%)	3 (2.9%)
L < M M < H	37	64	1 (2.7%)	3 (4.7%)	4 (4.0%)
M < H M > L	40	69	3 (7.5%)	2 (2.9%)	5 (4.6%)
H > M L < M	40	64	4 (10.0%)	2 (3.1%)	6 (5.8%)
M > L M < H	28	73	4 (14.3%)	2 (2.7%)	6 (5.9%)
M < H L < M	41	71	2 (4.9%)	5 (7.0%)	7 (6.2%)
M > L H > M	39	60	1 (2.6%)	6 (10.0%)	7 (7.1%)
Totals	297	547	15 (5.0%)	25 (4.6%)	40 (4.7%)

The error rates reported by other studies (where subjects read a problem typed on a card) were: De Soto et al[4] 39.2–61.7%, Clark[3] 6%, Potts[12] 5%, Mayer[10] 4–36%, Quinton et al[13] not given, Sternberg et al[16] 1.7–3.5%. A study where subjects heard a tape recoding of the problem[7] reported an error rate of 8–19%. In order to look for patterns in the errors made by subjects it is necessary to have a statistically significant sample of the errors made by them. Unfortunately, there were not enough incorrect answers to the `if` statement problem (Table 3) to enable any statistically significant analysis to be performed.

Possible techniques for producing a greater number of incorrect answers include: running the experiment for a longer period of time (it seems reasonable to assume that the number of errors will increase as the number questions answered increases), or by making the problem more difficult (e.g., using longer sounding identifiers).

5 General conclusions

It was hoped that the results of this experiment would provide some insight into subjects performance in handling short sequences of assignment and `if` statements. If the results of this experiment followed the pattern of behavior seen in other (non-software related) experiments, it would be possible to claim that the models of human cognition created to explain that behavior were also applicable here. The following summarizes the conclusions:

- *Assignment information held in working memory.* While there was some correlation between the duration of the spoken form of the identifiers appearing in assignment statements and subject performance, the content of long term memory also seems to play a significant role.

- *Performance differences in evaluating conditional expressions.* The form of relational expression had some impact on assignment recall performance (Figure 1). However, the operand orderings giving the *best* performance (i.e., lowest number of errors made when recalling assignment information) were not the same as those for which subject performed *best* (i.e., lowest number of incorrect answers to logic problem) in other studies.[3, 4, 7, 13, 15] There was insufficient error data (Figure 2) for any reliable statistical analysis of subject `if` statement evaluation performance to be carried out.

5.1 Where next?

While developers are often exhorted to think about the *meaningfulness* of identifiers, when creating new ones, the usability of identifiers within expressions and statements is rarely considered (apart, that is, from typing effort). More experiments need to be performed before it is possible to reliably draw any firm conclusions about the consequences of using different kinds of identifier spellings in assignment statements and on developer performance during source code comprehension. Other experiments might use a greater number of different character sequences (e.g., abbreviations, or identifiers containing two known words), randomize the order in which identifiers appear in the table of assignment answers, or use more commonly occurring character sequences. Other experiments might also use different filler tasks

Source code comprehension involves problem solving and developers are likely to use a variety of strategies to solve the problems that arise. The strategies used by developers can affect even such apparently simple tasks as remembering information about assignment statements. For instance, while some developers may choose to remember information about the identifiers appearing in an assignment using an encoding that involves their spoken form, other developers may use a different encoding (e.g., an abbreviated form of the identifier such as its' first letter, or the encoding of the semantics that the identifier represents). Any study of developer cognitive performance needs to ensure that the subjects taking part in an experiment are only using their cognitive resources in a way has been anticipated by the experimenter (even simple tasks such as counting have been found to require cognitive resources[1]).

The problems used in this study could be answered by subjects having insignificant amounts of experience in software development (e.g., undergraduate computer science students). It would be interesting to compare the performance of inexperienced subjects against that of subjects having a significant amount of experience. However, care needs to be taken when using inexperienced subjects to take into account the possibility of performance improvement through learning of the underlying coding problem itself.

6 Further reading

For a readable introduction to human reasoning see *Reasoning and thinking* by Ken Manktelow. *The Cognitive Animal* edited by M. Bekoff, C. Allen, and G. M. Burghardt contains 57 short, wide ranging, essays (of varying quality) on animal cognition.

6.1 Acknowledgments

References

At the end of each entry, the pages on which that entry is cited are listed in parentheses.

[1] V. Camos and P. Barrouillet. Adult counting is resource demanding. *British Journal of Psychology*, 95:19–30, 2004.

[2] H. H. Clark. Linguistic processes in deductive reasoning. *Psychological Review*, 76(4):387–404, 1969.

[3] C. B. De Soto, M. London, and S. Handel. Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology*, 2(4):513–521, 1965.

[4] J. S. B. T. Evans, J. L. Barston, and P. Pollard. On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, 11(3):295–306, 1983.

[5] A. S. Gilinsky and B. B. Judd. Working memory and bias in reasoning across the life span. *Psychology and Ageing*, 9(3):356–371, 1994.

[6] J. Huttenlocher. Constructing spatial images: A strategy in reasoning. *Psychological Review*, 75(6):550–560, 1968.

[7] D. M. Jones. Experimental data and scripts for short sequence of assignment statements study. `www.knosof.co.uk/cbook/accu04.html`, 2004.

[8] G. Leech, P. Rayson, and A. Wilson. *Word Frequencies in Written and Spoken English*. Pearson Education, 2001.

[9] R. E. Mayer. Qualitatively different encoding strategies for linear reasoning premises: Evidence for single association and distance theories. *Journal of Experimental Psychology: Human Learning and Memory*, 5(1):1–10, 1979.

[10] G. Paz-y-Miño C, A. B. Bond, A. C. Kamil, and R. P. Balda. Pinyon jays use transitive inference to predict social dominance. *Nature*, 430:778–781, Aug. 2004.

[11] G. R. Potts. Storing and retrieving information about ordering relationships. *Journal of Experimental Psychology*, 103(3):431–439, 1974.

[12] G. Quinton and B. J. Fellows. 'Perceptual' strategies in the solving of three-term series problems. *British Journal of Psychology*, 66:69–78, 1975.

[13] M. J. Roberts, D. J. Gilmore, and D. J. Wood. Individual differences and strategy selection in reasoning. *British Journal of Psychology*, 88:473–492, 1997.

[14] R. J. Sternberg. Representation and process in linear syllogistic reasoning. *Journal of Experimental Psychology: General*, 109(2):119–159, 1980.

[15] R. J. Sternberg and E. M. Weil. An aptitude $\times$ strategy interaction in linear syllogistic reasoning. *Journal of Educational Psychology*, 72(2):226–239, 1980.