Metamemory in Strategy Selection: Evidence for a Dissociation in Familiarity-Based Judgments between Semantics and Item Activation Strength

Alastair Mckenzie-Kerr (MckenzieKerrA@Cardiff.ac.uk)
School of Psychology, 64 Park Place,
Cardiff, CF10 3AT, UK

William J. Macken (Macken@Cardiff.ac.uk)
School of Psychology, Park Place, Cardiff University,
Cardiff, CF10 3AT, UK

Abstract

Two experiments investigated the metamemorial factors which influence rapid strategy selections (*retrieve* or *calculate*) in double-digit mental arithmetic problems. Using the GameShow paradigm (L. M. Reder & F. E. Ritter, 1992; C. D. Schunn et al., 1997) participants first predicted (within 850 ms) which strategy they would use to solve a sum and then solved the sum. Results revealed that predicted strategy selections were accurate, and that selection of the *retrieve* strategy was positively related to problem familiarity (L. M. Reder & F. E. Ritter, 1992; C. D. Schunn, et al., 1997), but not to the answer's familiarity (but see C. A. Lebiere & J. R. Anderson 1998; G. D. Logan, 2002). Furthermore, the semantic properties of a problem (e.g., whether both numbers are divisible by 5 or 10) appeared to dictate strategy selection. The evidence presented supports an account of strategy selection in which *retrieve* and *calculate* selections are determined by pre-retrieval indications of memorial strength derived from the familiarity of the problem (see also L. M. Reder & F. E. Ritter, 1992; C. D. Schunn et al., 1997) but also from semantic level representations of the problem's properties.

Keywords: Strategy Selection; Metacognition; Problem Solving; Familiarity; Semantic Memory; ACT-R

Introduction

Individuals use different strategies to solve different kinds of mathematical problems (Campbell & Timm, 2000; Hecht, 1999; LeFevre, Sadesky & Bisanz, 1996). When presented with a sum, such as 25 + 55 = ?, there are two distinct solution pathways available to the problem solver. If a direct fact retrieval from long term memory is used, the solution, 80, will be immediately available to the individual (Hecht, 1999; Lebiere & Anderson, 1998; Logan, 1988; Schunn et al., 1997; Siegler & Lemaire, 1997). This procedure is predominantly used by adults to solve *simple* sums such as 7 + 7 or 10 + 20. Alternatively, a calculation algorithm specific to the demands of the sum may be retrieved from long term memory and applied. Calculation algorithms are comprised of sub-processes drawn from a lexicon of procedures, likely to include direct fact retrieval, magnitude judgements, number comparison and rehearsal processes. For example, decomposition strategies break sums into manageable chunks which can be solved by directly retrieving known arithmetic facts (Hecht, 1999).

To date, most theoretical interest has focused upon the operation of direct retrieval and calculation algorithms as solution pathways in mental arithmetic (Lebiere & Anderson, 1998; Logan, 1988), rather than upon the factors that determine their deployment (Geary & Wiley, 1991; Reder & Lemaire, 1999; Siegler & Lemaire, 1997; Siegler & Stern, 1998).

One approach to strategy selection in arithmetic, characterised by models such as ACT-R 4.0 (Anderson & Lebiere, 1998, see also Siegler, 1987, 1988) and the Instance Theory of Attention and Memory (ITAM; Logan, 2002), negates the importance of identifying the problem characteristics that influence strategy deployment, proposing that solution strategies are automatically engaged upon sum presentation. Lebiere and Anderson's (1998) ACT-R conception of arithmetic processing suggests that direct fact retrieval commences immediately upon presentation of a sum. If this process fails to produce an answer of sufficient activation strength within a time limit specified by the task demands, then an appropriate calculation algorithm will be selected and applied. Alternatively, ITAM asserts that both retrieval and calculation procedures are engaged automatically in parallel, racing to produce an answer of sufficient activation strength to reach the threshold for output.

An alternative account, the Source Activation Confusion model (SAC), posits strategy selection as an implicit, rapid process operating prior to the engagement of direct fact retrieval and(/or) calculation algorithms (Reder & Ritter, 1992; Schunn et al., 1997). Strategy selection is determined by the strength of a preliminary Feeling of Knowing (pFoK) judgement, reflecting the belief that an item can be retrieved, based upon a very rapid search of memory. The strength of a pFoK is determined by a continuous index of activation reflecting the familiarity of the problem's terms. In the SAC model, if pFoK strength exceeds a predetermined threshold, the retrieval strategy will be selected (a metamemorial indicator of what is known within memory); alternatively the calculate strategy (indicative of what is not known in memory) will be selected.

Both experiments presented here utilise the GameShow methodology developed by Reder and Ritter (1992). Participants were presented with a sum and asked to indicate, if they had to answer it, whether they would retrieve the answer directly from long term memory or calculate the answer using an algorithm. A time limit of 850 ms prevented participants from solving the sum and using retrospect to identify the strategy they successfully used (Staszewski, 1988). In the second phase of each trial, participants were immediately requested to solve the sum. Using this methodology, Schunn et al. (1997) successfully manipulated the frequency of retrieve and calculate selections in double-digit multiplication sums by priming problem familiarity in a study phase.

We reasoned that in memory, arithmetic problems, similar to words (e.g., Hulme, Maughan & Brown, 1991), vary in their relative degree of familiarity. Accordingly, if strategy selections are made prior to the automatic engagement of solution procedures, problem familiarity may be used as a cue to strategy selection. The principle aim of both experiments was to demonstrate, using different types of stimuli, that rapid and accurate retrieve/calculate strategy selections can be made before problem solution procedures are completed (see also Reder & Ritter, 1992; Schunn et al., 1997). In this case, strategy selection may be determined by the familiarity of the problem, as the SAC model proposes, rather than the familiarity of solution procedures as both ACT-R and ITAM predict. The secondary aim in both experiments was to identify the problem features which determine the frequency of retrieve or calculate selections.

Experiment 1

Using the GameShow methodology participants were presented with double-digit addition problems which could only be solved by the calculate strategy. We tested three viable accounts of strategy selection, the ACT-R and ITAM accounts suggest that accurate strategy selection can only be made when solution procedures are allowed to complete. The time limit of 850 ms in the strategy selection phase precluded participants from completing the solution procedures and using retrospect to identify the successful strategy they used to answer the problem. Accordingly, if solution procedures were determining this process we would predict the frequency of retrieve/calculate selections to be at chance levels. The second account, which stands as an extension to the first, stipulates that it is possible that a measure derived from incomplete problem solution procedures may inform selection, specifically, the partial activation of the problem's answer (Schwartz & Metcalfe, 1992). Accordingly, problems were split into low and high answer familiarity conditions, effects of answer familiarity would indicate that a partial activation of the problem's answer guides strategy selection. Finally, the SAC model demonstrated that priming addend pairings and addend/operator pairings during a study phase increased the frequency of retrieve selections at test. Here we split problems into low and high sum familiarity conditions, predicting that for relatively high familiarity problems a greater frequency of retrieve selections would be evident, even though retrieve was the incorrect strategy selection in this experiment.

Methodology

Participants 24 undergraduates from the School of Psychology at Cardiff University were given course credit for their participation. All were native English speakers, reporting normal hearing and corrected or normal vision.

Materials & Design All problems were double-digit addition sums, each comprising two addends presented in the centre of the screen on one line, e.g., "31 + 37". The answer to each problem amounted to less than 100 and no problems were comprised of tied addends (e.g., "31 + 31"). Trials were arranged into different pseudo-random orders for each participant with 16 practice questions presented at the beginning of the experiment. Both addends in each problem were drawn from a sample ranging from 12 to 49 and were from the same decade class (e.g., "23 + 29"). Two variables were contrasted in a repeated measures design; *sum familiarity* and *answer familiarity*. Familiarity scores were derived from the number frequency measures developed by Gielen, Brysbaert and Dhondt (1991). Sum familiarity was obtained by adding together the familiarity ratings of the two addends in each problem. The second measure, answer familiarity, was purely derived from the familiarity rating of the answer. Each problem was ranked for sum familiarity and answer familiarity and assigned to high or low levels. Two repeated measures ANOVAs were conducted to ensure significant differences between the high and low levels of sum familiarity, $F(1, 15) = 105.34$, $MSE = 4142.89$, $p < 0.001$ and answer familiarity, $F(1, 15) = 412.63$, $MSE = 427.50$, $p < 0.001$.

Procedure Participants were instructed that they would be presented with a series of arithmetic problems, each trial comprising two phases. In phase 1, participants indicated how they would solve the presented problem (retrieve or calculate) within 850 ms (set to preclude direct retrieval of the answer; see also Reder & Ritter, 1992; Schunn et al., 1997; Staszewski, 1988). They were advised that this decision had to be made rapidly and that the time limit did not allow for much consideration of the task. If a strategy selection was made within 850 ms, participants immediately proceeded to phase 2. If not, phase 2 commenced 850 ms after the initial presentation of the problem. In phase 2, participants answered the problem as quickly and accurately as possible.

The procedure was self-paced, each trial commenced when the participant pressed the return key. A fixation mark ("X + X") was positioned in the centre of the screen, it flashed three times, each flash interleaved by a blank screen lasting for 850 ms. On what would have been the fourth flash of the fixation mark the problem appeared in its place.

At this point, participants were prompted to choose which strategy they would use to answer the problem; retrieve or calculate. To confirm a strategy selection, participants pressed a button on the keyboard marked "R" for retrieve or "C" for calculate. The "R" button was located on the z key of a standard qwerty keyboard and the "C" button on top of the m key.

Results & Discussion

Scoring Procedure Three measures were derived from participants' responses; a count of retrieve or calculate selections on each trial (*strategy selection),* the *strategy selection latency* measured from initial presentation of the problem until the retrieve/calculate selection and the *solution latency*. This was measured either from the entry of a strategy selection or expiry of the strategy selection time-window until the return key was pressed to confirm the response. For analysis, both latency measures were tagged to the preceding strategy selection, so on each trial a strategy selection latency and solution latency were tagged to either a retrieve or calculate strategy selection.

89.4% of strategy selections were made within the 850 ms timeframe. Only 4.7% of solution latencies were excluded from the analysis for breaching ±2 standard deviations from the mean. Incorrect scores accounted for only 7.4% of responses and did not vary systematically between conditions.

Strategy Selection As Table 1 illustrates, in all conditions the calculate strategy was chosen more frequently than retrieve (all ts > 5.5, all ps < 0.001), well above chance (50%) performance levels (all Fs > 30, all ps < 0.001). Separate 2 (sum familiarity; low v high) x 2 (answer familiarity; low v high) repeated measures ANOVAs analysed the frequency of retrieve and calculate selections. Participants were sensitive to sum familiarity in line with the predictions of the SAC model, fewer calculate selections were made in the high sum familiarity conditions than low sum familiarity condition, $F(1, 15) = 13.60$, $MSE = 9.10$, $p = 0.002$. Also, the frequency of retrieve and calculate selections was not influenced by answer familiarity; $F(1, 15) = 0.97$, $MSE = 7.76$, $p = 0.34$ and $F(1, 15) = 2.47$, $MSE = 7.74$, $p = 0.14$ respectively, suggesting that a partial activation of the answer did not influence strategy selection.

Selection Accuracy Table 1 illustrates the accuracy of strategy selection contrary to the predictions of ACT-R and ITAM which stipulate that completion of solution procedures and hindsight is required for accurate selection. Problems in which calculate was selected during phase 1 elicited longer solution latencies than when retrieve was selected at phase 1. However, this trend only reached significance in one condition; low sum familiarity/high answer familiarity, $t(1, 15) = 2.84$, $p = 0.01$. As none of the problems could be answered by the direct retrieval process this null effect indicates that the relatively low number of retrieve selections made were inaccurate, indicating that calculate, the correct strategy, was actually used to answer all problems.

Table 1: Mean strategy selection counts and solution latencies by condition.

	Low Answer Familiarity		High Answer Familiarity	
	Sum Familiarity		Sum Familiarity	
	Low	High	Low	High
Retrieve				
Strategy selected	19	28.1	21.1	35.2
Solution latency	4008	3981	3526	3487
Selection latency	534	601	572	583
Calculate				
Strategy selected	81	71.9	78.9	64.8
Solution latency	3998	4129	4170	3571
Selection latency	504	498	502	488

Note: Strategy selected represents the % of trials in which that strategy was chosen. Solution and strategy selection latencies are presented in ms.

Strategy Selection Latencies The pattern of results indicates that calculate selections were made more rapidly than retrieve selections. This reached significance in all (all ts > 2.85, all ps < 0.01) but the low sum familiarity/low answer familiarity condition ($t(1, 15) = 1.99$, $p = 0.07$) suggesting that separate thresholds determine retrieve and calculate selections. Alternatively, a single threshold account of strategy selection in which the duration of retrieve selections are protracted in an attempt to validate the selection may also account for this finding.

Furthermore, for retrieve selections, main effects of sum familiarity, $F(1, 15) = 4.81$, $MSE = 0.01$, $p = 0.04$ indicate that selections were made more rapidly for relatively familiar problems than unfamiliar problems. Calculate selection latencies were insensitive to sum familiarity $F(1, 15) = 2.69$, $MSE = 0.001$, $p = 0.12$ and both retrieve and calculate selection latencies were insensitive to answer familiarity, $F(1, 15) = 0.305$, $MSE = 0.004$, $p = 0.59$ and $F(1, 15) = 0.44$, $MSE = 0.002$, $p = 0.52$ respectively.

Returning to the three models of strategy selection, the first account, characterised by ACT-R and ITAM, suggested that completion of solution procedures and subsequent identification of the successful solution strategy would be required to ensure accurate strategy selection. Here however, within a time limit of 850 ms, the frequency of calculate strategy selections was significantly above chance levels, indicating that accurate selection could be made prior to the completion of solution procedures. If as the second account specifies, an *early read* or partial activation of the answer informs strategy selection then we would expect answer familiarity to influence the frequency of either calculate or retrieval selections. As only effects of sum familiarity were evident, the SAC model's prediction that strategy selection is determined by the familiarity of the problem provides the best fit to the data.

Experiment 2

Experiment 1 demonstrated that in line with the assumptions of the SAC model, problem familiarity can account for strategy selection. Further, strategy selection was shown to be immune to retrospective strategy identification taken upon completion of solution procedures or an early read of the answer. However, when applied to retrieve-calculate decisions in double-digit arithmetic, the SAC model failed to accommodate participants' strategy selections that did not appear to rely upon problem familiarity but the semantic properties of the problem. For example, Reder and Ritter (1992; Experiment 1) found that participants consistently chose the retrieve strategy for addition — as opposed to multiplication — problems in an attempt to beat the incentive system used in the methodology devised to stimulate rapid and accurate responses. In such instances, dictated by the task demands, participants apparently parsed the problem for semantic features such as operator types (i.e., + or ×) which were used to guide the selection process. Accordingly, Experiment 2, using the same methodology as Experiment 1, sought to establish whether retrieve/calculate selections are influenced by another semantic feature of problems, the common category membership of the addends. Three conditions of *sum type* were constructed; decades (e.g., 40 + 50), fives (e.g., 45 + 55) and mixed (e.g., 40 + 55). If a rapid analysis of the semantically classifiable problem features informs selection, in addition to a mechanism sensitive to sum familiarity, then effects of sum type would be expected.

Methodology

Participants 24 undergraduates from the School of Psychology at Cardiff University were given course credit for their participation. All were native English speakers, reporting normal hearing and corrected or normal vision.

Materials, Design & Procedure The problems presented in this experiment were drawn from a sample of decades ("20, 30, 40, 50, 60, 70") and fives numbers ("15, 25, 35, 45, 55, 65, 75"). Three conditions, each comprising 12 problems were constructed from purely decades, fives and a mix of decade and fives (e.g., "20 + 35"). To ensure that the familiarity of the problem or answer did not influence selection in the three sum type conditions, these factors were controlled in the construction of the stimuli, no significant differences between the sum type conditions existed in either measure (all *p*s > .05). The three sum types —*decades, mixed and fives*— were contrasted in a repeated measures design for analysis and the same procedure as detailed in Experiment 1 was used.

Results & Discussion

Scoring Procedure See Experiment 1 for details of the measures taken. 90.74% of strategy selections were made within the 850 ms timeframe and only 5.2% of solution latencies exceeded ± 3 standard deviations from the mean. Furthermore, only 4.85% of problems were answered incorrectly.

Strategy Selection As Table 2 illustrates, retrieve was selected more frequently than calculate in each sum type condition (all *t*s > 5.23, all *p*s < 0.001), significantly above chance level (50%) in each condition (all *t*s > 2.11, all *p*s < 0.05). Two repeated measures ANOVAs, demonstrated main effects of sum type for retrieve and calculate selections; $F(2, 10) = 13.66$, $MSE = 2.5$, $p = 0.001$ and $F(2, 10) = 9.04$, $MSE = 1.65$, $p = 0.006$ respectively. Planned pairwise comparisons revealed that retrieve was selected more frequently —hence calculate less frequently— in the decades condition than fives, also more frequently in the decades than mixed and mixed than fives (all *p*s < 0.05). Most revealing was the finding that 16 participants (out of 24) solely chose retrieve on each decade problem, 14 participants exclusively chose retrieve in the mixed condition and 10 participants in the fives condition. This suggests that in this task, strategy selection in the decades and mixed conditions was largely consistent with semantic categorisations of addends in each problem.

Table 2: Mean strategy selection counts and solution latencies by condition.

	Decades	Mixed	Fives
Retrieve			
Strategy selected	82.3	69.8	56.9
Solution latency	1405	1915	2799
Selection latency	532	568	616
Calculate			
Strategy selected	17.4	29.9	43.1
Solution latency	1950	2714	3323
Selection latency	515	530	549

Note: Strategy selected represents the % of trials in which that strategy was chosen. Solution and strategy selection latencies are presented in ms.

Selection Accuracy Confirming the accuracy of selection, solution latencies tagged to retrieve selections made at phase 1 were significantly shorter than those tagged to calculate selections in each sum type condition (all *t*s > 2.88, all *p*s < 0.02). Furthermore, two repeated measures ANOVAs found main effects of sum type for both calculate, $F(2, 10) = 32.52$, $MSE = 0.21$, $p < 0.001$ and retrieve selections, $F(2, 10) = 28.15$, $MSE = 0.19$, $p < 0.001$, suggesting that similar to the frequency of retrieve/calculate selections, the accuracy of strategy selection was also influenced by sum type.

Strategy Selection Latencies Similar to Experiment 1, the duration of retrieve selection latencies were consistently longer than calculate selection latencies. However, this effect only reached significant levels in the fives condition, $t(11) = 4.24$, $p = .001$, potentially due the relatively low frequency of calculate selections. Failure to replicate this effect precludes any further consideration at this point of whether the criteria for retrieve/calculate selections are set by a single or dual threshold mechanism.

Main effects of sum type upon the duration of retrieve selection latencies, $F(2, 10) = 32.98$, $MSE < .001$, $p < .001$ but not calculate selection latencies, $F(2, 10) = .77$, $MSE = .006$, $p = .49$ indicate that semantically classifiable features of the problem can influence the time taken to make a retrieve/calculate selection.

In sum, Experiment 2 demonstrated that semantically classifiable features of the problem, when available, can influence the frequency of retrieve/calculate selections. Replicating the effects of Experiment 1, strategy selections were accurate; solution latencies tagged to retrieve selections made at phase 1 were significantly shorter in duration than those tagged to calculate selections.

Conclusions

We believe the findings from the current series suggest a more sophisticated selection mechanism than previously conceived. In both experiments, contrary to the assumptions of ACT-R and ITAM, the accuracy of selection in problems normally solved by calculation (Experiment 1) and retrieval procedures (Experiment 2) was confirmed by the relation between strategy selections and solution latencies. Furthermore, an account of strategy selection determined by an early read of the answer (Schwartz & Metcalfe, 1992) was ruled out as null effects of answer familiarity were observed in Experiment 1. Findings from Experiment 1 support the intuitive notion that the more familiar a problem is within memory the more likely it is that an answer can be directly retrieved, befitting the assumptions of the SAC model. However, Experiment 2 illustrates that a rapid semantic level analysis of the numbers in a sum may also determine strategy selection contrary to the predictions of the SAC model. Previously, responses under speeded conditions, made prior to item retrieval and also in tests of recognition memory have been attributed purely to the monitoring of a continuous level of item activation (see Yonelinas, 2002 for review). Here however, a semantic level analysis of a problem’s terms has also been shown to influence rapid metamemorial responses.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C. & Qin, Y. (2004). An Integrated Theory of Mind. *Psychological Review, 111*(4), 1036-1060.

Belavkin, R. V. & Ritter, F. E. (2004). OPTIMIST: A New Conflict Resolution Algorithm for ACT-R. *Sixth International Conference on Cognitive Modelling*, Pittsburgh.

Campbell, J. I. D. & Timm, J. C. (2000). Adults' strategy choices for simple addition: Effects of retrieval interference. *Psychonomic Bulletin & Review, 7*(4), 692-699.

Gielen, I., Brysbaert, M. & Dhondt, A. (1991). The syllable-length effect in number processing is task dependant. *Perception & Psychophysics, 50*(5), 449-458.

Hecht, S. A. (1999). Individual solution processes while solving addition and multiplication math facts in adults. *Memory and Cognition, 27*(6), 1097-1107.

Hulme, C. S., Maughan, S. & Brown, D. G. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language 30*, 685-701.

Lebiere, C. A. & Anderson, J. R. (1998). Cognitive Arithmetic. In J. R. Anderson & C. Lebiere. (Eds.) *The Atomic Components of Thought*. (pp. 297-342). NJ; LEA.

LeFevre, J.-A., Sadesky, G. S. & Bisanz, J. (1996). Selection Procedures in Mental Addition: Reassessing the Problem Size Effect in Adults. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*(1), 2126-230.

Logan, G. D. (2002). An Instance Theory of Attention and Memory. *Psychological Review, 109*(2), 376-400.

Nelson, T. O. & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G.H. Bower (Ed). *The Psychology of Learning and Motivation*, 26, (pp. 125-173). NY: Academic Press.

Reder, L. M. & Ritter, F. E. (1992). What Determines Initial Feeling of Knowing? Familiarity With Question Terms, Not With the Answer. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*(3), 435-451.

Schunn, C. D., Reder, L. M., Nhouyvanisvong, A., Richards, D. R. & Stroffolino, P. J. (1997). To Calculate or Not to Calculate: A Source Activation Confusion Model of Problem Familiarity's Role in Strategy Selection. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*(1), 3-29.

Schwartz, B. L. & Metcalfe, J. (1992). Cue Familiarity but not Target Retrievability Enhances Feeling-of-Knowing Judgements. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*(5), 1074-1083.

Siegler, R. S. & Lemaire, P. (1997). Older and Younger Adults' Strategy Choices in Multiplication: Testing Predictions of ASCM Using the Choice/No-Choice Method. *Journal of Experimental Psychology: General 126,* (1), 71-92.

Staszewski, J. J. (1988). Skilled Memory and Expert Mental Calculation. In M. T. H. Chi, R. Glaser & M. J. Farr (Ed.s) *The Nature of Expertise.* New Jersey, LEA.

Yonelinas, A. P. (2002). The Nature of Recollections and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language, 46*, 441-517.