THE CHIMP CHALLENGE: VISUAL WORKING MEMORY IN CHIMPS AND HUMANS

SEÁN ROBERTS

Language and Cognition Department
Max Planck Institute for Psycholinguistics, The Netherlands
sean.roberts@mpi.nl

JUSTIN QUILLINAN

Language Evolution and Computation Research Unit
School of Philosophy, Psychology & Language Sciences, The University of Edinburgh
justin@ling.ed.ac.uk

Matsuzawa (2012) presented work at Evolang demonstrating the working memory abilities of chimpanzees. (Inoue & Matsuzawa, 2007) found that chimpanzees can correctly remember the location of 9 randomly arranged numerals displayed for 210ms - shorter than an average human eye saccade. Humans, however, perform poorly at this task. Matsuzawa suggests a semantic link hypothesis: while chimps have good visual, eidetic memory, humans are good at symbolic associations. The extra information in the semantic, linguistic links that humans possess increase the load on working memory and make this task difficult for them. We were interested to see if a wider search could find humans that matched the performance of the chimpanzees. We created an online version of the experiment and challenged people to play. We also attempted to run a non-semantic version of the task to see if this made the task easier. We found that, while humans can perform better than Inoue and Matsuzawa (2007) suggest, chimpanzees can perform better still. We also found no evidence to support the semantic link hypothesis.

The limited-hold memory task (Inoue & Matsuzawa, 2007, hereafter I&M) is a game where the participant sees a random array of numerals displayed on a screen in a random configuration. The numerals are masked after a short latency. The participant must then press the masking symbols in the order suggested by the symbols that they cover (see figure 1). If they pressed all the displayed numbers in the correct order (e.g. ordinal ascending) then they get a reward. If they press one out of sequence, the trial ends and no reward is given. In I&M, two chimps and 9 humans were tested. One of the chimps, Ayumu, completed about 80% of trials correctly, regardless of latency. Humans and the other chimpanzee, Ai, performed worse at lower latencies, scoring between 20-40% at 210ms. The differences between the chimpanzees might be due to age (Ayumu was younger) or, interestingly, because Ai had been language-trained, while Ayumu had not. Mat-

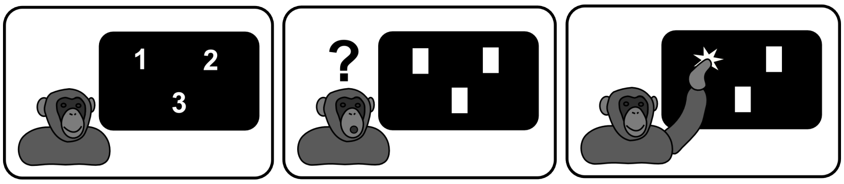

Figure 1. The experimental setup: Symbols are presented then masked. The participant attempts to select the masked symbols in their correct sequence.

suzawa suggested that the semantic links possessed by humans puts extra pressure on working memory, making the task difficult. Matsuzawa speculated that this kind of visual ability is adaptive for chimpanzees, who regularly fight with other groups and need to keep track of where many attackers are at once, while ranking them by threat level.

210ms is at the limit of the time a human eye can move from from central to peripheral stimuli (Bartz, 1962), which means that a human should not be able to fixate all numbers on the screen before they are masked. Humans can produce 'express saccades' at 100ms if their visual field is blank before the stimuli appears rather than if there is a central fixation point displayed up until the stimulus appears (Fischer & Ramsperger, 1984, but see Kingstone & Klein, 1993 for an alternative hypothesis). This is known as the 'gap effect' and the conditions for express saccades are essentially met in the limited hold task. However, chimpanzees do not exhibit a 'gap effect', performing equally well in both conditions (Kano, Hirata, Call, & Tomonaga, 2011), so the success of chimpanzees at the limited hold task cannot be because of this. Kano and Tomonaga (2011) demonstrate that while both species have similar saccade latencies in the gap condition, chimpanzees make more fixations per second and have shorter fixation durations than humans, making them more effective at visual scanning. Kano and Tomonaga suggest that humans and chimpanzees have different strategies for resolving competition between central and peripheral stimuli.

Silberberg and Kearns (2009) suggested that the chimpanzees were benefitting from more practice with the task. The results in I&M came from 500 trials for chimpanzees but only 50 trials for humans. After extensive training, the human participants in Silberberg and Kearns (2009) were performing comparably with Ayumu for 5 numerals. However, there are reports of chimpanzees performing correctly 80% of the time with 8 numerals at 210ms (Matsuzawa, 2009).

1. Eidetic Memory

The study of the ability to recall intricate visual images or eidetic memory has a long history. (Sperling, 1960) discovered that the number of letters that participants could correctly recall seeing in an array presented for a short amount of time did not differ as a function of the latency. Later studies of 'informational persistence' showed that the identity of a symbol shown for a short amount of time could be held between 500ms, while the location of that symbol could only be held for 150-300ms (Loftus & Irwin, 1998; Luck & Hollingworth, 2008). It is the second type of memory that appears to be weak in humans in the limited-hold memory task.

However, there are individuals with very good eidetic memory. Up to 15% of children have good eidetic memory[a], but this declines in adulthood (Haber & Haber, 1988). There are some adult individuals with extraordinary eidetic memory. Stephen Wiltshire can draw extremely complex and accurate pictures of cityscapes from memory (and without sketching). Wiltshire would presumably perform similarly to Ayumu on the limited-hold memory task. Interestingly, Wiltshire's language development was delayed. However, there are other adult individuals with strong eidetic memory who also have good phonological memory (e.g. 'Elizabeth', see Stromeyer & Psotka, 1970). There are also human individuals with highly superior autobiographical memory, but normal visual reproduction memory (LePort et al., 2012)[b].

With regards to the semantic hypothesis, (Lupyan & Spivey, 2008) find that adding semantic information actually improves visual search. Participants had to identify a target symbol amidst similar distractor symbols. Pointing out that the symbols could be seen as '2's and '5's rotated 90 degrees improved participants' performance on the task. However, this task measured visual identification, not on the recall of location. An alternative (but not mutually exclusive) hypothesis is that humans have evolved an advanced phonological working memory, trade-off against a visual working memory (see Hurford, 2010). A further hypothesis is that Ayumu has developed synaesthetic associations between number and colour (Humphrey, 2012). Synaesthetic associations in humans usually involve learned sequences of stimuli and there is evidence that having colour associations can aid recall (Gross, Neargarder, Caldwell-Harris, & Cronin-Golomb, 2011).

2. The QHImp Qhallenge

We adapted this task to run online, using a similar layout and look to the original experiments (coded in javascript and HTML). You can see the implementation at http://www.correlation-machine.com/qhimp/. In order to attract participants, we presented the experiment as a game called the *QHImp Qhallenge* (Quick-Hold Improvement Challenge). In order to give participants practice with the game there were three 'modes' in which to play. 'Arcade mode' presented increasingly

[a] 2-15% of children aged 6-10 shown a detailed picture that was then removed could accurately describe the picture and did so in the present tense, suggesting that the visual perception was still accessible to them.

[b] See amazing autobiographical memory http://tiny.cc/qjsk8w

difficult trials starting with a single symbol at a long latency, then decreasing the latency while increasing the number of symbols. 'Challenge mode' started with 9 symbols for 3 seconds and decreased the latency slowly to 210ms. 'Chimp mode' presented 9 symbols for 210 milliseconds. To proceed to the next level of difficulty, participants had to choose all symbols in the current trial in the correct order. If participants made a wrong choice, the trial would end and they would 'lose a life'. Once all their 'lives' were lost, the game ended and they would have to start again. Participants gained an arbitrary number of points for passing each level and there was a 'leaderboard' which displayed the players with the highest scores.

The stimuli were always sequential and started from the first in the set. Therefore, if there were 4 numerals on the screen, these would always be 1, 2, 3 and 4. In I&M, the stimuli set was sampled randomly, so there were possible gaps in the sequence. As a reviewer pointed out, this makes the task more difficult, since participants can't always infer the identity of a symbol by exclusion. We're currently addressing this issue (see section 4).

2.1. *Stimuli*

There were five sets of stimuli. First, as in I&M, numerals from 1 to 9. We also attempted to choose sets of symbols without semantic links. We used letters arranged alphabetically from a to i, shades of colour from red to white to green[c], a set of arbitrary symbols (presented in the same arbitrary order for all participants) and directional arrows arranged clockwise (see figure 2). We asked participants to choose a player name and recorded the participant's responses. Because this program would be run on a range of computers, operating systems and web browsers, the actual latency time could not be guaranteed to match the intended latency time. Therefore, the actual latency time was recorded.

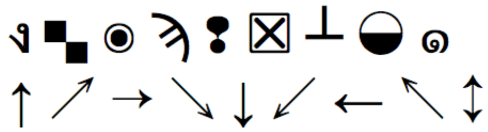

Figure 2. Two sets of non-semantic symbols used in the experiment.

[c]The range had to be fairly broad to support easily recognisable differences between the shades

Table 1: Results of a linear regression testing how various factors predict performance.

Factor	Estimate	t-value	p-value	
Number of previous trials	1.08E-04	42.836	<0.00001	***
Numerals displayed	-4.73E-02	-17.098	<0.00001	***
Saccade Distance	-2.54E-03	-5.892	<0.00001	***
Latency	2.37E-05	3.004	0.00267	**
Order (top to bottom)	-2.06E-02	-1.86	0.06295	.
Order (left to right)	-2.76E-03	-0.25	0.80266	

2.2. *Participants*

Participants were recruited online using blogs and twitter. No monetary incentive was offered. Around 170 participants took part in our experiment (identified by names they chose themselves), completing a median of 24 trials each (range between 1 and 3382). A total of around 16,500 trials were recorded over one month. Although there were ways participants could cheat, since we could not observe them directly, the incentive to do so was low and we found no evidence of cheating after checking the average trial completion time for each participant.

3. Results

Although the results of this experiment are difficult to compare to the original experiment, the first question was whether the humans in the current experiment were any better than those in I&M. Figure 3 shows the results for all participants

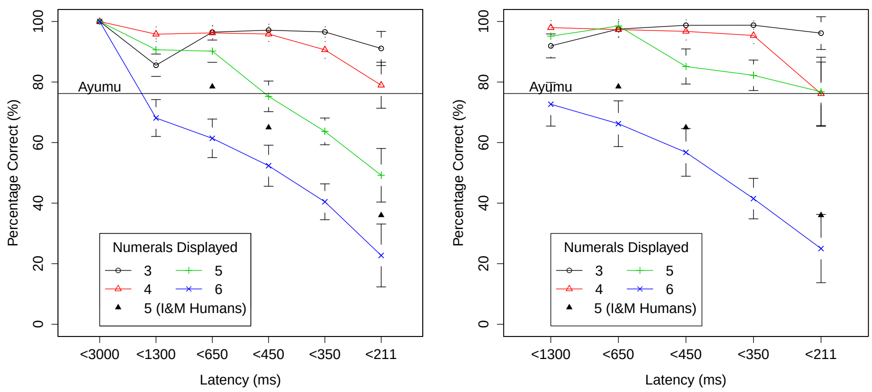

Figure 3. Results for the experiment with 95% confidence intervals for numeral stimuli for all participants (left) and for the best 66 participants (right). The solid horizontal line indicates Ayumu's performance and the solid triangles indicate the performance of the humans on 5 numerals from I&M.

for a range of latencies and number of numerals displayed. While humans appear to be performing slightly better than the human participants of I&M, they are clearly not performing as well as Ayumu at 210 ms. However, there were a large individual differences in performance. Looking at players who got over 75% of their trials correct on five numbers (at all latencies), a different picture emerges (figure 3). 66 participants have a mean within 5% of Ayumu's performance for 5 numerals shown for 210 ms or less. This result is qualitatively different from that of I&M. The best overall player at 210ms latency got 80% of his trials correct on 6 numerals (5 trials) and 60% of his trials correct on 7 numerals (5 trials). The best player for 9 numerals at 210ms only got 18% of their trials correct (174 trials). Ayumu, whose performance is unaffected by latency up to 8 numerals, still appears to be better than humans at this task.

Out of 2347 attempts at choosing 9 numerals at 210ms latency, 75 were successful (3%). Although humans' performance is much lower than Ayumu's, humans are choosing the correct numeral at a rate greater than chance up to 7 numerals (figure 4). Interestingly, humans choose the first numeral correctly about 80% of the time, which is comparable to Ayumu's performance on all 9 numerals. This suggests that humans can *identify* symbols at this latency (as in Sperling, 1960), but cannot retain the *location* of these symbols for very long.

Finally, we used some post-hoc tests to look at various factors that make the task hard for humans. The proportion of numbers guessed correctly in a trial was entered as the dependent variable into a linear regression model. The independent variables were the number of numerals displayed, the actual latency time, the game 'mode' and how many previous trials the participant had played. We also

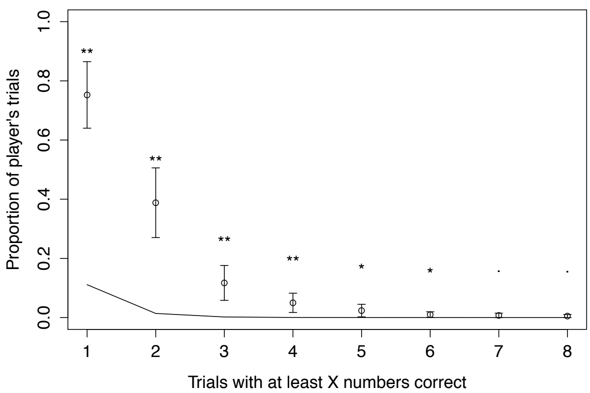

Figure 4. Results for 9 numerals displayed for 210ms or less. The solid line represents the level of success predicted by chance. Note that once the player chooses 8 numbers correct, there is only one numeral left to choose from. Symbols above means give significance above chance. **: $p < 0.001$, *: $p < 0.01$, $\cdot$: $p \approx 0.05$.

measured how difficult the layout of the numerals might have been to remember. The saccade distance measured the cumulative distance between each numeral in order, assuming that a shorter saccade trajectory would allow more numerals to come within fixation. Also, numerals presented in order from top to bottom or from left to right might make the task easier (measured using Damerau Levenshtein distance).

The results are shown in table 1. As expected, the number of numerals displayed and how much practice the participant had (number of previous trials) were significant factors. However, the latency predicted less of the variation in performance than the saccade distance, suggesting that humans were attempting to fixate the individual numerals. How ordered the numerals were from top to bottom was only a marginally significant predictor.

3.1. *Non-Semantic Symbols*

We compared the proportion of symbols guessed correctly in each trial over the different stimuli types. Contrary to the predictions of the semantic link hypothesis, 10% fewer symbols were guessed correctly per trial on average from the non-semantic stimuli compared with the numeral stimuli ($t=19.04$, $p < 0.0001$, linear regression controlling for number of symbols displayed, latency, mode, number of previous trials played, saccade distance and ordering, model accounting for 54% of the variance, see figure 5). No participant performed significantly better with

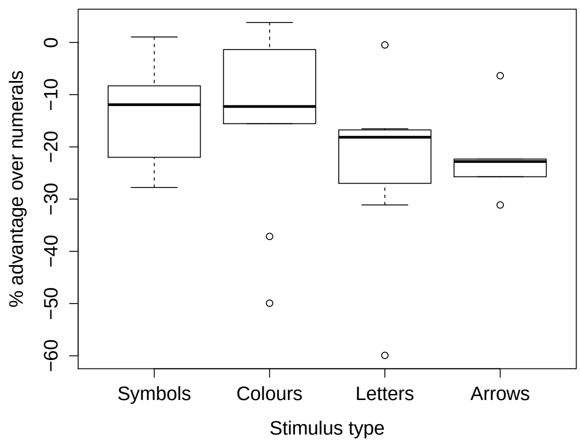

Figure 5. Percentage difference in performance within trials between numeral and non-semantic stimuli for 5 symbols displayed.

any of the non-semantic stimuli than the numerals. However, the results are not entirely clear-cut. Two participants did show a small, non-significant improvement with colour stimuli. In light of the synaesthetic hypothesis (Humphrey, 2012) it may be useful to test this further. Also, one participant also showed a small, non-significant improvement with the arbitrary symbols. The poorer performance could be due to the non-semantic stimuli being less familiar than numerals or perhaps the participants forged semantic links between non-semantic stimuli during training.

4. Future Research

In collaboration with the Living Links Research Centre in the Edinburgh Zoo, we have installed a version of this experiment in the Living Links Science Exploration Zone, an area for visitors to the zoo to interact with and learn about different research projects. This installation addresses several issues with the previous experiment. Firstly, as in I&M, we use non-sequential stimuli, where the numbers displayed will be drawn from a random sample from one to nine. We are also gathering data on the age of the participants. Since eidetic memory in humans declines in adulthood (Haber & Haber, 1988), and performance of chimps on this task is said to decrease with age (Humphrey, 2012), it will be useful to determine how age affects performance in humans. Finally, this installation consists of a single machine, so the dimensions and layout of the screen will be standard across participants, and timings can be measured with greater accuracy.

5. Conclusion

This study found evidence that humans can perform better than suggested in I&M in the limited-hold memory task. However, human performance is still below that of chimpanzees. This difference appears to stem from an inability to keep the *location* of symbols in working memory. We found no evidence to support the hypothesis that semantic links with the symbols were interfering with the task.

The methodology harnessed the power of online crowdsourcing of participants. We hope that this experiment demonstrates that empirical results can be brought to bear on research questions rapidly and cost-effectively, without the attendant ethical issues with utilising online labour markets. The results are available to download at: http://www.correlation-machine.com/qhimp/data/.

References

Bartz, A. (1962). Eyemovement latency, duration, and response time as a function of angular displacement. *Journal of Experimental Psychology*, *64*(3), 318.

Fischer, B., & Ramsperger, E. (1984). Human express saccades: extremely short reaction times of goal directed eye movements. *Experimental Brain Research*, *57*(1), 191–195.

Gross, V. C., Neargarder, S., Caldwell-Harris, C. L., & Cronin-Golomb, A. (2011). Superior encoding enhances recall in color-graphemic synesthesia. *Perception-London*, *40*(2), 196.

Haber, R., & Haber, L. (1988). Eidetic imagery as a cognitive skill. In L. Obler & D. Fein (Eds.), (Vol. The exceptional brain: The neuropsychology of talent and special skills, p. 218-241). New York: Guildford Press.

Humphrey, N. (2012). This chimp will kick your ass at memory games - but how the hell does he do it? *Trends in cognitive sciences*, *16*(7), 353–355.

Hurford, J. (2010). *The origins of grammar.* Oxford University Press.

Inoue, S., & Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Current Biology*, *17*(23), R1004–R1005.

Kano, F., Hirata, S., Call, J., & Tomonaga, M. (2011). The visual strategy specific to humans among hominids: A study using the gap–overlap paradigm. *Vision research*.

Kano, F., & Tomonaga, M. (2011). Species difference in the timing of gaze movement between chimpanzees and humans. *Animal cognition*, 1–14.

Kingstone, A., & Klein, R. (1993). What are human express saccades? *Attention, Perception, & Psychophysics*, *54*(2), 260–273.

LePort, A., Mattfeld, A., Dickinson-Anson, H., Fallon, J., Stark, C., Kruggel, F., Cahill, L., & McGaugh, J. (2012). Behavioral and neuroanatomical investigation of highly superior autobiographical memory (hsam). *Neurobiology of Learning and Memory*.

Loftus, G., & Irwin, D. (1998). On the relations among different measures of visible and informational persistence. *Cognitive Psychology*, *35*(2), 135–199.

Luck, S., & Hollingworth, A. (2008). *Visual memory.* Oxford University Press, USA.

Lupyan, G., & Spivey, M. (2008). Perceptual processing is facilitated by ascribing meaning to novel stimuli. *Current Biology*, *18*(10), R410–R412.

Matsuzawa, T. (2009). Symbolic representation of number in chimpanzees. *Current opinion in neurobiology*, *19*(1), 92–98.

Matsuzawa, T. (2012). Outgroup: The study of chimpanzees to know the human mind. In T. C. Phillips, M. Tamariz, E. A. Cartmill, & J. R. Hurford (Eds.), (Vol. The Evolution of Language: Proceedings of the 9th International Conference (EVOLANG9), p. 227-233). World Scientific.

Silberberg, A., & Kearns, D. (2009). Memory for the order of briefly presented numerals in humans as a function of practice. *Animal cognition*, *12*(2), 405–407.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological monographs: General and applied*, *74*(11), 1.

Stromeyer, C., & Psotka, J. (1970). The detailed texture of eidetic images. *Nature*, *225*(5230), 346.