Foraging Connections: Optimal Foraging in Letter Fluency

Kimberly Arjune (ksarjune@syr.edu)
Jeffrey C. Zemla (jczemla@syr.edu)
Department of Psychology
Syracuse University, Syracuse, NY

Abstract

The letter fluency task is the timed listing of words that begin with a specific letter (e.g., words starting with T). Participants often list words in phonologically related clusters (e.g., *tank*, *task*, *tap*) and occasionally switch clusters (e.g., *tap*, *thud*). This process has been likened to patch switching in animal foraging. Optimal performance requires switching clusters in a manner that maximizes the rate of retrieving words, known as the marginal value theorem. Previous work has found evidence for this in semantic fluency. The current study tests whether people adhere to the marginal value theorem in letter fluency and whether executive functioning is associated with optimal performance. Three letter cues (T, N, and J) and one semantic cue (animals) were administered. Results are consistent with optimal search in N and J, but not T or animals. These findings provide mixed support that people search optimally during letter fluency.

Keywords: letter fluency; marginal value theorem; executive functioning; semantic fluency

Introduction

Verbal fluency is a timed word retrieval task under specific conditions. There are two types of verbal fluency tasks: semantic fluency, which is listing items from a semantic category such as animals (Bousfield & Sedgewick, 1944), and letter fluency, which is listing words that start with a specific letter (Borkowski, Benton, & Spreen, 1967). In both of these tasks, participants tend to cluster similar words together (Troyer, Moscovitch, & Winocur, 1997). For example, one might list *lion*, *tiger*, *cat* (all felines) when listing animals or *trip, trap, trunk* (all *tr*- words) when listing words that start with the letter T. Clusters are interrupted by occasional switches to dissimilar words.

This search process has been likened to animals foraging for food, which is often modeled as search over clustered patches of resources (Stephens & Krebs, 1986). As resources in a cluster are depleted, animals face diminishing returns and eventually leave the cluster in search of a new one with greater resources. The optimal search solution (optimal foraging) is one that maximizes the overall rate of encountering resources, which is achieved by switching clusters when the rate within a cluster drops to the average rate across all clusters. This is known as the marginal value theorem (Charnov, 1976).

Prior work has examined the marginal value theorem in the context of semantic fluency (Abbott, Austerweil, & Griffiths, 2015; Hills, Jones & Todd, 2012; Montez, Thompson, & Kello, 2015; Zemla, Gooding, & Austerweil, 2023). The retrieval rate (i.e., the rate of retrieving words from memory) is measured with inter-item response times (IRTs). Data are largely consistent with the marginal value theorem: within-cluster IRTs increase until they approach (but do not exceed) the average IRT across all clusters, while between-cluster IRTs are larger than within-cluster IRTs, reflecting the cost of searching for a new cluster. This has been taken as evidence of optimal search in the semantic fluency task. We apply some of the same analyses to letter fluency, a related task but with key differences.

Evidence of a double dissociation between semantic and letter fluency impairments in patient populations indicates that the two tasks rely, at least partially, on different cognitive processes and brain regions. For example, individuals with Alzheimer's disease show greater impairments on semantic than letter fluency (Henry, Crawford, & Phillips, 2004), while people with aphasia show greater impairments on letter than semantic fluency (Bose et al., 2022).

Executive functioning is thought to be involved in both tasks, evidenced by frontal lobe patients who show equal impairments at both tasks (Henry & Crawford, 2004). Measures of executive functioning like operation span are also correlated with both fluency types (Shao et al., 2014). However, a concurrent task (divided attention) impairs performance on letter fluency more than semantic fluency (Moscovitch, 1994), suggesting executive functioning may have a greater role in letter than semantic fluency. Marko et al. (2023) use factor analysis to provide additional evidence for this, and suggest that executive functioning is responsible for suppressing automatic semantic associations in letter fluency. Older adults tend to generate larger clusters (i.e., lower switch rate) than younger adults in letter fluency but not semantic fluency (Troyer et al., 1997).[1] This is notable because aging is associated with deficits in executive functioning (Hasher, & Zacks, 1988). Troyer et al. (1997) contend that switches involve more executive control processes than non-switches, though control processes likely play a role in non-switches as well (Mayr & Kliegl, 2000).

These findings provide some support for the differential cognitive processes involved in semantic and letter fluency. The specifics are often debated, and there are counterexamples to the above references. However, there is broad agreement that there are differences in the underlying

---

[1] Traditionally, switching is measured by counting the number of switches in a fluency list (Troyer et al., 1997). This measure has been criticized (Hills et al., 2013; Mayr, 2002), in part because it is highly correlated with the total number of responses. In contrast, cluster size may be more informative because it is nearly proportional to the *rate* of switching.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

cognitive processes of the two tasks and that executive functioning is an important component of search.

The current study examines whether participants adhere to the marginal value theorem in letter fluency (letters T, N, and J). Our primary analyses parallel those by Hills et al. (2012), who found evidence of optimal search in semantic fluency. Our results were consistent with the marginal value theorem for some letters (N and J). However for T fluency, and when collapsed across all three letters, our results indicated that participants spent too long in a cluster before switching. Surprisingly, we found this to be the case in semantic fluency as well, in contrast to prior work.

In addition, we tested whether a measure of executive functioning predicts adherence to the marginal value theorem. We measured executive functioning using the operation span task (Turner & Engle, 1989). We predicted that individuals high in executive functioning would switch more optimally, but found no support for this.

Method

Participants

One hundred and eight students were recruited from an undergraduate subject pool in exchange for course credit. Data from 89 participants were analyzed (19 were excluded for not pressing Enter during the task, leading to incomplete collection of response times). Fifty eight participants were female, 28 male, 1 other, and 2 unknown. Their age range was 18-27 (M = 19.19, SD = 1.34).

Materials and Methods

Data collection was done in a lab setting. Participants began with a practice trial of the letter fluency task (letter F) for thirty seconds. Then, participants completed letter fluency trials for T, N, and J (three minutes each), in that order. We chose these letters because we expected them to vary in the number of responses, with T being easy, N being moderate, and J being hard (Borkowski et al., 1967). Letters were chosen to manipulate the difficulty (availability) of words across trials. Borkowski et al. (1967) found these letters to differ in both fluency performance and in word frequency from a corpus of 4.5 million words from popular magazine articles (Thorndike & Lorge, 1944). Participants were instructed not to list proper nouns (e.g., Tom, Tampa) or morphological variants of the same word (e.g., talk, talking). Participants typed responses into a text box, and responses faded from screen after the participant hit Enter. Next, participants completed a three minute semantic fluency task (animals).

After the fluency tasks, participants completed an operation span (OSPAN) task to measure executive functioning, following the scoring protocol of Unsworth et al. (2005) with minor adjustments. We used OSPAN as our measure of executive functioning because Shao et al. (2014) found a strong correlation between OSPAN and the number of responses listed in both letter and semantic fluency. On each trial, participants saw a sequence of letters (one letter per screen) with basic arithmetic problems presented between letters. Participants were asked to remember the sequence of letters (in order) while solving the math problems. The sequences increased from length three (three letters, three math problems) to seven (seven letters, seven math problems). Each sequence length was repeated three times for a total of 15 trials. No feedback was provided during the task. Following Unsworth et al. (2005), participants who scored less than 85% accuracy on the math problems were excluded from the executive functioning analyses (ten participants). The OSPAN was scored by summing all of the perfectly recalled sets (e.g., if they recalled all 3 letters in a 3 set, all 4 letters in a 4 set, and 3 of the letters in a 5 set, their score would be 3 + 4 + 0 = 7). Finally, participants completed a demographic questionnaire to record age, gender, race, and whether they have been diagnosed with dyslexia or another language impairment (five participants).

Data analysis

Rule-breaking responses (i.e., proper nouns, different forms of a word) were not excluded. The reason for this was to avoid the creation of fictive clusters.[2] However, we used an automated word stemmer (Hunspell, 2015) to approximate the number of rule-break responses and found a low number of them that did not vary substantially across letters (0.81% for T, 1.06% for N, 1.14% for J). Phonological clusters were defined as two or more consecutive responses with the same first two letters. Typically, this is one of several crtiteria for identifying cluster switches (e.g., Tröster et al., 1998). Clusters in the semantic fluency task were defined as consecutive responses in which each transition is between two animals of the same subcategory (e.g., *cat* and *lion* are both felines). These were determined using SNAFU (Zemla, Cao, Mueller, & Austerweil, 2020), a tool for segmenting semantic fluency data. Responses were spell-corrected for semantic fluency to ensure that animals were properly categorized. Each response was assigned positions relative to the previous cluster switch (e.g., 1 indicates the first item in a cluster) and relative to the next cluster (e.g., -1 indicates the last item in a cluster).

Results were analyzed on a group basis. For each individual and each cluster position, a ratio was calculated indicating the mean IRT for that position divided by the mean IRT across all positions. For each cluster position, a one-sample *t*-test was used to assess whether the IRTs were larger than average (ratio significantly greater than one) or smaller than average (ratio significantly less than one).

We followed the procedure of Hills et al. (2012) to test for optimal foraging. First, we expected that IRTs would

---

[2] For example, suppose a participant lists *train*, *Tennessee*, *track*. The removal of *Tennessee* results in one cluster (*train*, *track*) instead of three.

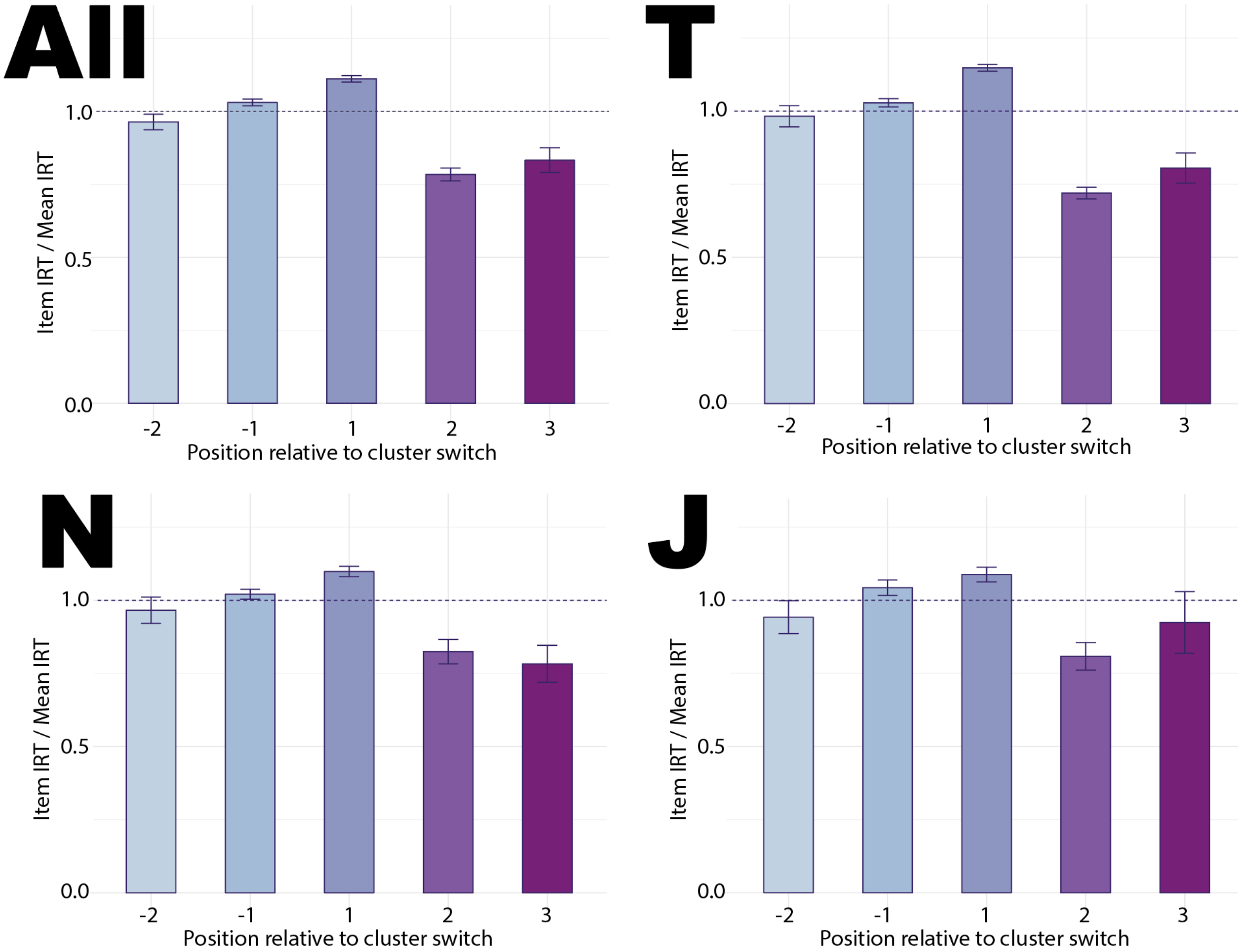

Figure 1. Plots indicate whether the mean IRT for each cluster position was greater or less than the average across all cluster positions, separated by letter. All (upper left) collapses across letters T, N, and J. Error bars represent the standard error of the mean.

monotonically increase within a cluster. Second, we expected that IRTs for the last item in a cluster would not significantly differ from the average. This pattern is consistent with the marginal value theorem which prescribes that people should leave a cluster when the marginal rate of return within a cluster is equal to the average rate of return across all clusters. Third, we expected that the first item in a cluster would significantly exceed the average rate of return, as this IRT includes time searching for a new cluster. Fourth, we expected that the second item in a cluster would be significantly smaller than the average response time, as the rate of retrieval should be high when entering a new cluster. We repeated this process for each fluency category (T, N, J, and semantic). In addition, we repeated the process collapsing across all letter fluency trials using a weighted average, so as to not bias towards any particular letter (an unweighted mean biases the average towards letters with more responses).

We tested whether participants who were more consistent with the marginal value theorem listed more responses. We used the absolute difference between the last item IRT (position -1) and the average IRT as an individual measure of consistency with the marginal value theorem. We used this measure (hereafter referred to as the deviation score) to predict the total number of responses. We expected deviation scores to be negatively correlated with the number of items listed.

We also used deviation scores to predict OSPAN. A negative association was expected, with higher OSPAN scores predicting lower deviation scores. That is, we expected participants high in executive functioning to be more consistent with the marginal value theorem. We conducted this analysis first collapsing across letters and then for each category individually.

Results

Participants listed more words starting with T (M = 32.75) than N (M = 20.73), $t(88) = 19.48$, $p < .001$, and more words starting with N than J (M = 16.44), $t(88) = 24.43$, $p < .001$. Participants listed more animals (M = 36.28) than they did with the letter T, $t(88) = 3.85$, $p <$ 001.

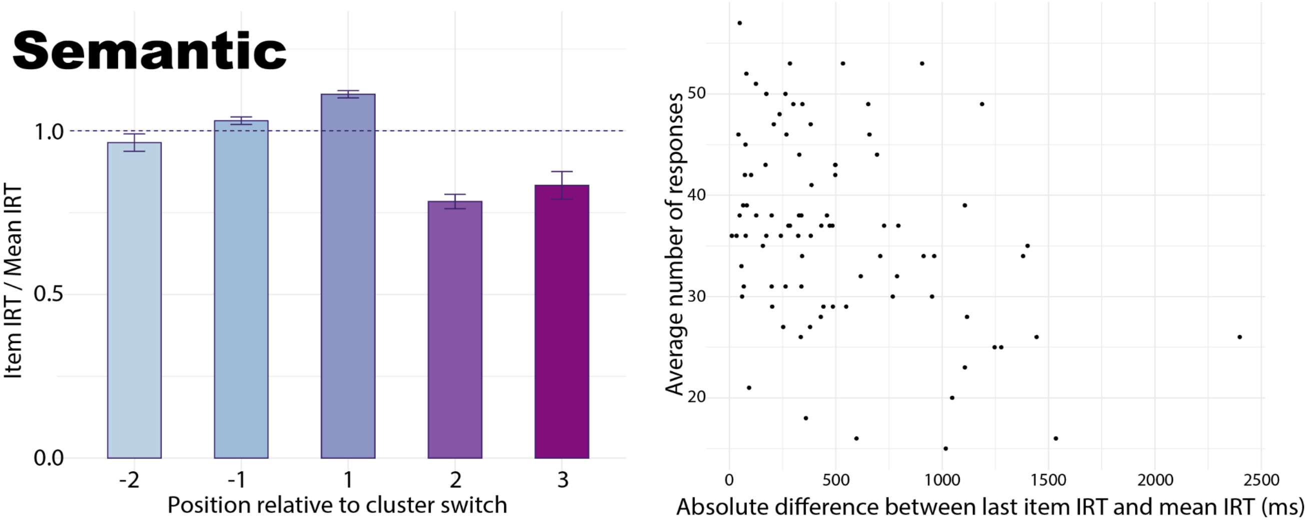

Figure 2. (Left) Plot indicates whether the mean IRT for each cluster position was greater or less than the average across all cluster positions. Error bars represent the standard error of the mean. (Right) On average, participants with a smaller absolute difference between last item IRT and mean IRT (deviation score) listed more responses

Optimal foraging

All three letters (as well as collapsing across letters) qualitatively showed a similar pattern of increasing IRTs within a cluster (see Figure 1).

When collapsing across letters (Figure 1, top left), the last cluster position (-1) had significantly larger IRTs than average, $t(88) = 2.45$, $p = .01$. The first cluster position (1) also had significantly larger IRTs than average, $t(88) = 7.48$, $p < .001$. The second cluster position had significantly smaller IRTS than average, $t(88) = -8.25$, $p < .001$.

For letter T (Figure 1, top right), the last cluster position (-1) had significantly larger IRTs than average, $t(88) = 2.00$, $p = .04$. The first cluster position (1) also had significantly larger IRTs than average, $t(88) = 12.77$, $p < .001$. The second cluster position had significantly smaller IRTs than average, $t(88) = -13.94$, $p < .001$.

For letter N (Figure 1, bottom left), there was no significant difference between the last cluster position (-1) and the average IRT, $t(88) = 1.24$, $p = .21$. The first cluster position (1) had significantly larger IRTs than average, $t(88) = 5.58$, $p < .001$. The second cluster position had significantly smaller IRTs than average, $t(87) = -4.21$, $p < .001$.[3]

For letter J (Figure 1, bottom right), there was no significant difference between the last cluster position (-1) and the average IRT, $t(88) = 1.62$, $p = .11$. The first cluster position (1) had significantly larger IRTs than average, $t(88) = 3.5$, $p < .001$. The second cluster position had significantly smaller IRTs than average, $t(86) = -4.07$, $p < .001$.[3]

For semantic fluency (see Figure 2, left), the last cluster position (-1) had significantly larger IRTs than average, $t(88) = 2.69$, $p = .008$. The first cluster position (1) also had significantly larger IRTs than average, $t(88) = 11.92$, $p < .001$. The second cluster position had significantly smaller IRTs than average, $t(88) = -13.23$, $p < .001$.

Deviation scores were significantly correlated with the number of responses generated in all categories (see Figure 2 for semantic fluency, Figure 3 for letter fluency). Collapsed across letters, $r(87) = -.36$, $p < .001$. For T, $r(87) = -.32$, $p = .002$. For N, $r(87) = -.20$, $p = .05$. For J, $r(87) = -.22$, $p = .05$. For semantic fluency, $r(87) = -.38$, $p < .001$.

Executive Functioning

Deviation scores were not significantly associated with OSPAN scores in any of the conditions. When collapsed across letters, $t(76) = .83$, $p = .40$. For T, $t(76) = 1.4$, $p = .16$. For N, $t(76) = 1.6$, $p = .1$. For J, $t(76) = .96$, $p = .34$. For semantic fluency, $t(76) = .308$, $p = .76$.

Discussion

The results are partially (though not entirely) consistent with our prediction that the timing of cluster switching would abide by the marginal value theorem. For all letters, IRTs tended to increase at each position within a cluster. In addition, the first item in a cluster had significantly larger IRTs than average, and the second item in a cluster had significantly smaller IRTs than average. For two letters (N and J), the last item in a cluster was approximately equal to (i.e., not significantly different than) the average IRT. This is consistent with the marginal value theorem, which states that participants should switch clusters when the rate of retrieval within a cluster is equal to the average rate across all clusters. However for semantic fluency, letter T, and when collapsed across letters, participants spent

---

[3] Degrees of freedom vary because some participants did not generate clusters with more than one word.

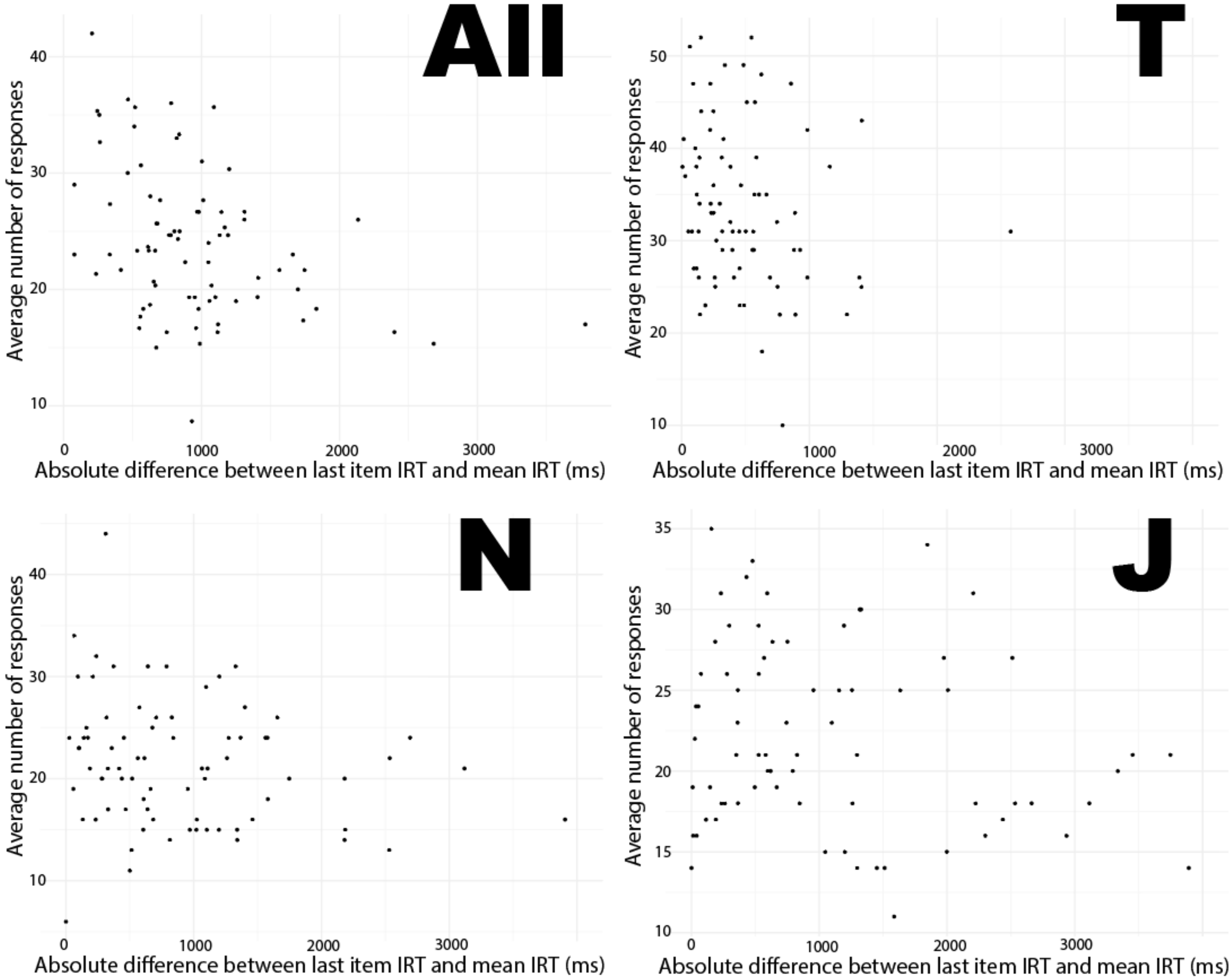

Figure 3. On average, participants with a smaller absolute difference between last item IRT and mean IRT (deviation score) listed more responses. All (upper left) collapses across letters T, N, and J. Plots are truncated at 4000ms on the x-axis.

significantly longer in a cluster before switching than is prescribed by the marginal value theorem. Deviation scores were significantly correlated with the number of responses, highlighting that consistency with the marginal value theorem is beneficial to performance, as individuals who can effectively move between clusters are able to retrieve more words at a faster rate. These results are surprising given that many studies have found that performance in the semantic (animal) fluency task is consistent with the marginal value theorem (Abbott et al., 2015; Hills et al., 2012; Zemla et al., 2023). However, semantic similarity has been measured several ways in the past, including the use of distributional semantic models to demarcate clusters (Hills et al., 2012). Our method might be less sensitive to cluster switches. Moreover, in other domains, it is not uncommon to find that participants violate the marginal value theorem by spending more time than is optimal. For example, Hutchinson, Wilke, and Todd (2008) found that participants in a virtual fishing game spent too long fishing within a pond before switching to a new pond. Zacksenhouse, Bogacz, and Holmes (2010) examined speed-accuracy tradeoffs in a two-alternative forced choice task, and found that participants spent too long on trials, and as a result did not maximize reward rate.

One explanation for the discrepancy between letter types relates to practice effects. On the first letter fluency trial (letter T), participants spent too long in a cluster before switching. However on subsequent letter fluency trials (N and J), participants were better calibrated. Prior work suggests that participants may become more optimal with practice. For example, Zemla et al. (2023) conducted a repeated semantic fluency task (three animal fluency trials with filler questionnaires in between). They found that while performance was mostly in line with the marginal value theorem, adherence increased with each trial. In a two-alternative forced choice task, Balci et al. (2011) found that participants do not initially maximize reward rate, but learn to do so over time. They attribute this to uncertainty in time estimation, a factor that could apply to verbal fluency as well. Most participants demonstrate improved performance on repeated testing of the letter fluency test, indicating that participants adapt to the task and learn effective strategies (Bird et al., 2004; Harrison et al., 2000). However, participants also spent too long in clusters during the semantic fluency task, which was the very last fluency trial. Practice effects alone would suggest that participants should be closest to optimal on this task, though differences between the two tasks may limit the transfer of practice effects.

A second explanation relates to the varying level of difficulties across the conditions. Adherence to the marginal value theorem was found in the conditions where the fewest number of responses were listed (i.e., the hardest conditions; N and J) while behavior was suboptimal in the conditions where the most number of responses were listed (T and semantic). Note that since the task has a fixed time length, the number of responses listed is inversely proportional to the average IRT. Lundin et al. (2023) examined the neural correlates of switching in semantic and letter fluency using fMRI. They found evidence for a ramping neural signal that precedes a switch and correlates with IRTs in multiple brain regions. It is possible that when IRTs are small, participants are capable of generating a larger number of responses before this signal reaches a threshold and initiates a switch. In contrast, when IRTs are large, this ramping signal may reflect the urgency of switching. This explanation offers an algorithmic and implementational account of the difference between easier and harder categories, though it still implies that participants fail to switch optimally on easier trials. In work on modeling speed-accuracy tradeoffs, Holmes & Cohen (2014) found that participants may appear suboptimal when assessed by the criterion of reward rate maximization in part because they are also trying to minimize the cost of cognitive control.

The use of null hypothesis significance testing may also be problematic. In particular, the marginal value theorem states that participants should switch clusters when the marginal rate of retrieval within a cluster is equal to the average rate of retrieval across all clusters. This is tested using a one-sample *t*-test on the last cluster position, and a null result is predicted. This is similar to the analyses of Hills et al. (2012) and Abbott, Austerweil, and Griffiths (2015), differing only in that they conduct this test on data from individual participants as opposed to the sample as a whole. Common statistical wisdom suggests that null results should not be used as evidence for the null hypothesis. The analysis also treats the distinction as binary: participants either leave a cluster at the appropriate time, or they do not. However when we instead consider the magnitude of deviation from optimal policy, as in Figure 2, we see that participants who are closer to optimal list more responses. In fact, correlations were strongest for semantic fluency and T, despite (or perhaps because) participants stayed in clusters too long for those categories.

Limitations and Future Directions

Our analyses offer a first look at whether cluster switching in the letter fluency task is consistent with the marginal value theorem. However, the analyses are more limited in scope and differ somewhat from previous work. For example, prior work has proposed a process model to explain this behavior as search over a semantic representation (Hills et al., 2012; Abbott, Austerweil, & Griffiths, 2015). In the Introduction, we contend that letter and semantic fluency rely on different cognitive processes, but we do not propose a computational account of these differences. Some process models of the task explicitly use different retrieval cues (word frequency, semantic similarity, phonological similarity) at cluster switch points (e.g., Kumar, 2022). Our work ignores these complexities, which could have influenced our results. Similarly, we rely on a crude measure of segmenting clusters (using the first two letters), which does not account for overall phonological similarity (e.g., rhymes; Koren, Kofman, & Berger, 2005) or the influence of non-phonological similarity (e.g., semantic or syntactic) in letter fluency (Schwartz et al., 2003). Other studies have used automated tools such as VFClust (Ryan et al., 2013) for this purpose (Lundin et al., 2023), or used multiple criteria along with the first two letters (Tröster et al., 1998; Kosmidis et al., 2004). Without a better computational account of the task, it is difficult to assess whether participants are truly suboptimal, or whether various constraints of the task affect how one should define optimality.

We did not find evidence that OSPAN was correlated with adherence to the marginal value theorem. We used the OSPAN as a measure of executive functioning because prior work has found it to be correlated with verbal fluency performance (Shao et al., 2014). However, executive functioning is a multifaceted construct, encompassing updating, shifting, and inhibition (Miyake et al., 2000), and OSPAN may not be the ideal measure related to switching in verbal fluency. While OSPAN emphasizes short-term memory and updating (Shao et al., 2014), participants have little need to explicitly store information in working memory. Others have argued that inhibition (Hirshorn & Thompson-Schill 2006; Marko et al., 2023) and shifting (Amunts et al., 2020) play key roles in verbal fluency. For a better measurement of the executive functions used in letter fluency, future studies could focus on tasks that emphasize these aspects of executive functioning, such as the Stroop or cued task switching.

Our study examined the relationship between the marginal value theorem with lexical retrieval using a letter fluency task. While some conditions aligned with the marginal value theorem (N, J), other conditions (T, semantic) did not show this same pattern. Along with this, there was not a relationship between OSPAN and deviation score. These results differed from our expectations based on prior literature, and future studies may reconcile these inconsistencies.

References

Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging. *Psychological Review*, *122*(3), 558-569.

Amunts, J., Camilleri, J. A., Eickhoff, S. B., Heim, S., & Weis, S. (2020). Executive functions predict verbal fluency scores in healthy participants. *Scientific Reports*, *10*(1), 11141.

Balci, F., Simen, P., Niyogi, R., Saxe, A., Hughes, J. A., Holmes, P., & Cohen, J. D. (2011). Acquisition of decision making criteria: Reward rate ultimately beats accuracy. *Attention, Perception, & Psychophysics*, 73, 640–657.

Bird, C. M., Papadopoulou, K., Ricciardelli, P., Rossor, M. N., & Cipolotti, L. (2004). Monitoring cognitive changes: Psychometric properties of six cognitive tests. *British Journal of Clinical Psychology*, *43*(2), 197-210.

Borkowski, J. G., Benton, A. L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia*, *5*(2), 135-140.

Bose, A., Patra, A., Antoniou, G. E., Stickland, R. C., & Belke, E. (2022). Verbal fluency difficulties in aphasia: A combination of lexical and executive control deficits. *International Journal of Language & Communication Disorders*, *57*(3), 593-614.

Bousfield, W. A., & Sedgewick, C. H. W. (1944). An analysis of sequences of restricted associative responses. *The Journal of General Psychology*, *30*(2), 149-165.

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, *9*(2), 129-136.

Harrison, J. E., Buxton, P., Husain, M., & Wise, R. (2000). Short test of semantic and phonological fluency: Normal performance, validity and test-retest reliability. *British Journal of Clinical Psychology*, *39*(2), 181-191.

Hasher, L. & Zacks, R. T. (1988). Working Memory, comprehension, and aging: A review and a new view. Psychology of Learning and Motivation, *22*, 193-225

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology*, *18*(2), 284-295.

Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. Neuropsychologia, *42*(9), 1212-1222.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory.*Psychological Review*, *119*(2), 431.

Hills, T. T., Mata, R., Wilke, A., & Samanez-Larkin, G. R. (2013). Mechanisms of age-related decline in memory search across the adult life span. *Developmental Psychology*, *49*(12), 2396-2404.

Hirshorn, E. A., & Thompson-Schill, S. L. (2006). Role of the left inferior frontal gyrus in covert word retrieval: Neural correlates of switching during verbal fluency. *Neuropsychologia*, *44*(12), 2547-2557.

Holmes, P., & Cohen, J. D. (2014). Optimality and some of its discontents: Successes and shortcomings of existing models for binary decisions. *Topics in Cognitive Science*, *6*, 258–278.

Hunspell (2015). Home Page. Retrieved April 25, 2025, from http://hunspell.github.io.

Hutchinson, J. M., Wilke, A., & Todd, P. M. (2008). Patch leaving in humans: can a generalist adapt its rules to dispersal of items across patches? *Animal Behaviour*, *75*(4), 1331-1349.

Koren, R., Kofman, O., & Berger, A. (2005). Analysis of word clustering in verbal fluency of school-aged children. *Archives of Clinical Neuropsychology*, *20*(8), 1087-1104.

Kosmidis, M. H., Vlahou, C. H., Panagiotaki, P., & Kiosseoglou, G. (2004). The verbal fluency task in the Greek population: Normative data, and clustering and switching strategies. *Journal of the International Neuropsychological Society*, *10*(2), 164-172.

Kumar, A. A., Lundin, N. B., & Jones, M. N. (2022). Mouse-mole-vole: The inconspicuous benefit of phonology during retrieval from semantic memory. In *Proceedings of the 44th Annual Meeting of he Cognitive Science Society*.

Lundin, N. B., Brown, J. W., Johns, B. T., Jones, M. N., Purcell, J. R., Hetrick, W. P., ... & Todd, P. M. (2023). Neural evidence of switch processes during semantic and phonetic foraging in human memory. *Proceedings of the National Academy of Sciences*, *120*(42), e2312462120.

Marko, M., Michalko, D., Dragašek, J., Vančová, Z., Jarčušková, D., & Riečanský, I. (2023). Assessment of automatic and controlled retrieval using verbal fluency tasks. *Assessment*, *30*(7), 2198-2211.

Mayr, U., & Kliegl, R. (2000). Complex semantic processing in old age: Does it stay or does it go? *Psychology and Aging*, *15*(1), 29-43.

Mayr, U. (2002). On the dissociation between clustering and switching in verbal fluency: Comment on Troyer, Moscovitch, Winocur, Alexander and Stuss. *Neuropsychologia*, *40*(5), 562-566.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49-100.

Montez, P., Thompson, G., & Kello, C. T. (2015). The role of semantic clustering in optimal memory foraging. *Cognitive Science*, *39*(8), 1925-1939.

Moscovitch, M. (1994). Cognitive resources and dual-task interference effects at retrieval in normal people: The role of the frontal lobes and medial temporal cortex. Neuropsychology, *8*(4), 524-534.

Ryan, J. O., Pakhomov, S., Marino, S., Bernick, C., & Banks, S. (2013, August). Computerized analysis of a verbal fluency test. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 884-889).

Schwartz, S., Baldo, J., Graves, R. E., & Brugger, P. (2003). Pervasive influence of semantics in letter and category fluency: A multidimensional approach. *Brain and Language*, *87*(3), 400-411.

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*, 772.

Stephens, D. W., & Krebs, J. R. (1986). Foraging theory (Vol. 1). Princeton University press.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words.* New York, NY: Teachers College, Columbia University

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138-146.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127-154.

Tröster, A. I., Fields, J. A., Testa, J. A., Paul, R. H., Blanco, C. R., Hames, K. A., ... & Beatty, W. W. (1998). Cortical and subcortical influences on clustering and switching in the performance of verbal fluency tasks. *Neuropsychologia*, *36*(4), 295-304.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498-505.

Venkatagiri, H. S., & Levis, J. M. (2007). Phonological awareness and speech comprehensibility: An exploratory study. *Language Awareness*, *16*(4), 263-277.

Zacksenhouse, M., Bogacz, R., & Holmes, P. (2010). Robust versus optimal strategies for two-alternative forced choice tasks. *Journal of Mathematical Psychology*, *54*, 230–246.

Zemla, J. C., Cao, K., Mueller, K. D., & Austerweil, J. L. (2020). SNAFU: The semantic network and fluency utility. *Behavior Research Methods*, *52*, 1681-1699.

Zemla, J. C., Gooding, D. C., & Austerweil, J. L. (2023). Evidence for optimal semantic search throughout adulthood. *Scientific Reports*, *13*(1), 22528.