Retrieval Mechanisms in Sentence Comprehension

by

Jordan Ashley Whitlock

B.A., Linguistics
Cornell University, 2011

Submitted to the Harvard-MIT Health Sciences and Technology Program
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Health Sciences and Technology

at the

Massachusetts Institute of Technology

June 2014

Signature of Author .... Signature redacted ..
Harvard-MIT Health Sciences and Technology
May 16, 2014

Certified by .... Signature redacted
David Caplan, MD, PhD
Professor of Neurology, Harvard Medical School
Thesis Supervisor

Accepted by .... Signature redacted
Emery N. Brown, MD, PhD
Director, Harvard-MIT Program in Health Sciences and Technology
Professor of Computational Neuroscience and Health Sciences and Technology

RETREIVAL MECHANISMS IN SENTENCE COMPREHENSION

by

Jordan Ashley Whitlock

Submitted to the Harvard-MIT Health Sciences and Technology Program
on May 16, 2014 in partial fulfillment of the
requirements for the Degree of Master of Science in
Health Sciences and Technology

ABSTRACT

This work investigates the nature of the memory mechanisms utilized in language comprehension. Through the use of the Speed Accuracy Tradeoff (SAT) paradigm (Wickelgren, 1977), healthy young adults were studied for the use of parallel or serial search mechanisms to understand syntactically complex sentences with multiple embeddings. Systematically designed sentence stimuli tested whether the relevant memory mechanism differs when reanalysis is required. Results indicated that sentence length and syntactic ambiguity affected overall accuracy of sentence comprehension. The rate in which information was retrieved did not vary for most sentence types, but may have been affected by length in one type of sentence (ambiguous "early closure" sentences). The data support a parallel, content-addressable retrieval mechanism for information in most sentences but may provide evidence for serial search in ambiguous sentences that require complex syntactic reanalysis.

Thesis Supervisor: David Caplan, MD, PhD

Title: Professor of Neurology, Harvard Medical School

I. Acknowledgments:

I would like to thank a few of the many people who have supported me through the journey of graduate school. First, I express my appreciation and thanks to my thesis supervisor, David Caplan, for guidance throughout the data collection, analysis, and writing stages of this project. Thank you also to Julie Van Dyke at Haskins Laboratory for support with the SAT methodology, use of analysis code, and assistance with data analysis. To Adam Ostrowski and Will Evans, thank you for your support and patience during data collection. To my classmates at both MIT and MGH, thank you for keeping me focused during busy semesters. Thank you to my family for your full, enthusiastic support of all of my academic and career decisions. Finally, to Charles and Auggie, my deepest gratitude for both your aid with figures and support throughout the entire graduate school process.

The project described was supported by Grant Number T32 DC00038 from the National Institute on Deafness and Other Communication Disorders (NIDCD). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders (NIDCD) or NIH.

II. Introduction

This study investigates the nature of the memory mechanisms utilized in language comprehension. Through the use of the Speed Accuracy Tradeoff (SAT) paradigm (Wickelgren, 1977), neurologically normal young adults were studied for the use of parallel or serial search mechanisms to understand syntactically complex sentences with multiple embeddings. Systematically designed sentence stimuli tested whether the relevant memory mechanism differs when reanalysis is required. Results of the study add to the understanding of the relationship between memory and language comprehension in healthy young adults, and may later be used as a comparison against individuals with disordered memory or language comprehension.

Language understanding begins with a complex sensory signal and ends in the successful transmission of information. Though the process is seemingly automatic, it is incredibly complex. Sensory input must be broken down into units that are then put together in a structure that provides the comprehender with meaning. Throughout the comprehension process, there is constant monitoring and feedback. This work concerns comprehension components related to the assignment of sentence structure and meaning, monitoring, and feedback.

The experiment reported here investigates Caplan and Waters' model of the comprehension process. Caplan and Waters posited a bipartite architecture for language comprehension consisting of a skilled parser and a controlled processor (2013) . The skilled parser quickly and automatically assigns syntactic structure to incoming constituents. Based on monitoring and feedback, the controlled processor acts in a deliberate, slower manner to provide structure and meaning. Meaning is derived from the parser unless it fails, in which case the controlled processor is utilized.

In most cases, as a sentence is presented, the first assignment of syntactic structure proves to be adequate. In some instances, however, the initial structure and meaning assignment produces a meaning that is in some way unacceptable. These are the cases in which, according to the Caplan and Waters bipartite model, the analysis of the slower-working processor takes over. Ambiguities are one situation in which a skilled parser may fail to operate adequately. Other such structures are center embeddings, which are hard for listeners to comprehend (Lewis, Vasishth, & Van Dyke, 2006) . Increasing the amount of information in a sentence by adding words or embedding clauses also increases difficulty in comprehension (McElree, Foraker, & Dyer, 2003) .

Waters and Caplan argued that the parser and processor utilize two distinct memory mechanisms: a parallel, content-addressable search, and a serial search, respectively (2013) . In parallel search, all items represented in memory are accessed simultaneously (McElree et al., 2003). The search process ends when the target item is identified or the probability that the item is not present in accessible memory reaches a cutoff point. Due to the simultaneous access of all items in the search set, the speed at which an individual retrieves accessible information is expected to be equal. In serial search, each item accessible in memory is searched individually until a target is identified or fails to be (McElree & Dosher, 1993) . Search may be forward (from the beginning of utterance), backward (most recent item, followed by the preceding item, etc.), or use of other algorithms (Henson, 1998). Speed of retrieval is therefore expected to slow as search set size increases. For both parallel and serial search, the probability of accessing the correct information is based on the size of the search space. As the number of items in the search space increases, the accuracy of retrieving the intended item decreases.

Evidence for the role of parallel search is found for retrieval of information from both lists and sentences. Varying serial position of target items and list size for retrieval of target items in lists changed only overall accuracy for the task, but not time course, suggesting a parallel search (McElree & Dosher, 1989). Increasing the distance between an NP and its missing argument by embedding additional clauses did not change the rate of item retrieval, providing evidence for parallel search regardless of length (McElree, 2000). Changing the location of interfering information within a sentence resulted in accuracy difference, but equivalent dynamics of accuracy increases over time (Martin & McElree, 2009) .

The mechanism of serial search is primarily supported by evidence from studies of list retrieval. When asked to judge recency of a target item in a list, subjects demonstrated accuracy dynamics consistent with serial search: speed of information retrieval varied with regard to the size of the search set (McElree & Dosher, 1993) . Thus far, only one study of sentence comprehension has produced evidence that may support serial search for comprehension of sentences. McElree and colleagues found a difference in the time course of accuracy for sentences with two embedded object relative clauses as compared to all other sentences types they tested (McElree et al., 2003). They claimed, however, that this was not evidence for a serial search process because rate did not systematically slow based on sentence complexity and length.

In summary, there is evidence for both parallel and serial search in retrieval of linguistic information, with parallel begin tied to determining the presence of items, and serial to retrieving temporal information. A substantial body of literature supports content-addressable mechanisms for sentence comprehension, but difficulty of sentences has not been significantly increased.

Waters and Caplan assigned parallel search to the parser, and serial search to the controlled processor. Because all items are accessed simultaneously in parallel search, it is fast. This speed is characteristic of the automatic parser. The incremental, slow nature of serial search aligns more closely with the controlled processor component of the model.

The order of items is important for comprehension of languages such as English, so some degree of serial search may be expected in some capacity in language comprehension. Furthermore, comprehension of adequately complex or ambiguous linguistic information is expected to fail during parser-based comprehension, and may require analysis by the controlled processor. The expectation, therefore, is that evidence for a serial search mechanism may be found in comprehension of highly complex or ambiguous sentences.

The work reported here aims to determine whether the time course and accuracy of retrieval of items in syntactically complex, ambiguous sentences demonstrate the use of parallel or serial search mechanisms, and in turn, whether the bipartite comprehension model of Caplan and Waters is supported empirically.

If evidence of serial search mechanisms emerges for one or more sentence types, then this would support the bipartite language comprehension theory of an automatic parser and skilled processor. If for all sentence types, regardless of syntactic complexity or ambiguity, subjects demonstrate a response pattern consistent with parallel search, then this will call into question the validity of the bipartite model.

III. Methods

This work utilized a multiple-response Speed-Accuracy Tradeoff (SAT) paradigm with a sentence acceptability task. The speed-accuracy tradeoff concept claims that the faster an individual responds to a stimulus, the less likely s/he is to be accurate. Given additional processing time, accuracy increases. Thus, there is an effective "trade-off" between speed and accuracy of responses. This effect is robust across fields and has been investigated in sensory modalities such as vision and olfaction (Rinberg, Koulakov, & Gelperin, 2006; Sutter & Graham, 1995) . The Speed-Accuracy Tradeoff (SAT) experimental paradigm takes advantage of this effect to measure the time course of information retrieval (Wickelgren, 1977).

According to the notion of the speed-accuracy tradeoff, response accuracy varies as a function of time. An experimental paradigm such as reaction time captures a single speed and accuracy data point. In his argument for the use of the SAT paradigm over reaction time experiments, Wickelgren claimed that data collected in reaction time experiments are frequently from the portion of the speed-accuracy curve where accuracy is not yet asymptotic.

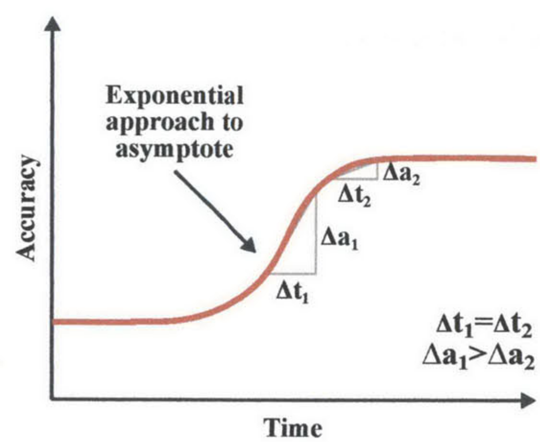


Figure 1: SAT versus RT: small, equal change in time leads to unequal changes in accuracy

A small difference in accuracy, therefore, would correspond to a large difference in reaction time (see Figure 1). If data were taken from lower-accuracy regions of the curve, small accuracy differences would correspond to small reaction time differences. Because it is not possible to ascertain where in the decision-making process a reaction time occurs, reaction time experiments are limited in the conclusions that they can make (Wickelgren, 1977). A paradigm that collects responses at controlled latencies (such as the Speed-Accuracy Tradeoff) is advantageous because a more complete estimate of the SAT function is created.

Early experiments using the SAT procedure utilized a single-response method (e.g. see Dosher et al., 1989; McElree & Dosher, 1989; McElree et al., 2003). Stimulus presentation was followed by a single auditory cue, presented at one of several controlled latencies. Each stimulus type was tested multiple times, with each timing interval tested at least once. Data for each stimulus type was then collapsed to create a speed and accuracy response curve. Because each stimulus presentation collected a single data point, the cued response required a large number of trials and was thus time-intensive and limited the number of subjects to be included in each study.

The multiple-response method (Martin & McElree, 2009; Van Dyke & McElree, 2011) reduces the number of trials necessary in each experiment, therefore reducing time required to collect data. In a multiple-response implementation of the SAT paradigm, each stimulus presentation is followed by multiple auditory cues for subject response. The auditory cue, a short tone, is presented at equal intervals (e.g., every 300ms), and subjects respond with a key press. Timing and accuracy data are collected across a several second period for each stimulus. The advantage of the multiple-response method is that a larger number of data points can be collected in one trial, thus reducing the number of necessary trials and shortening total subject participation time. Additionally, accuracy curves are collected within a single trial rather than being extrapolated from multiple trials, thus increasing the validity of patterns in speed and accuracy for responses.

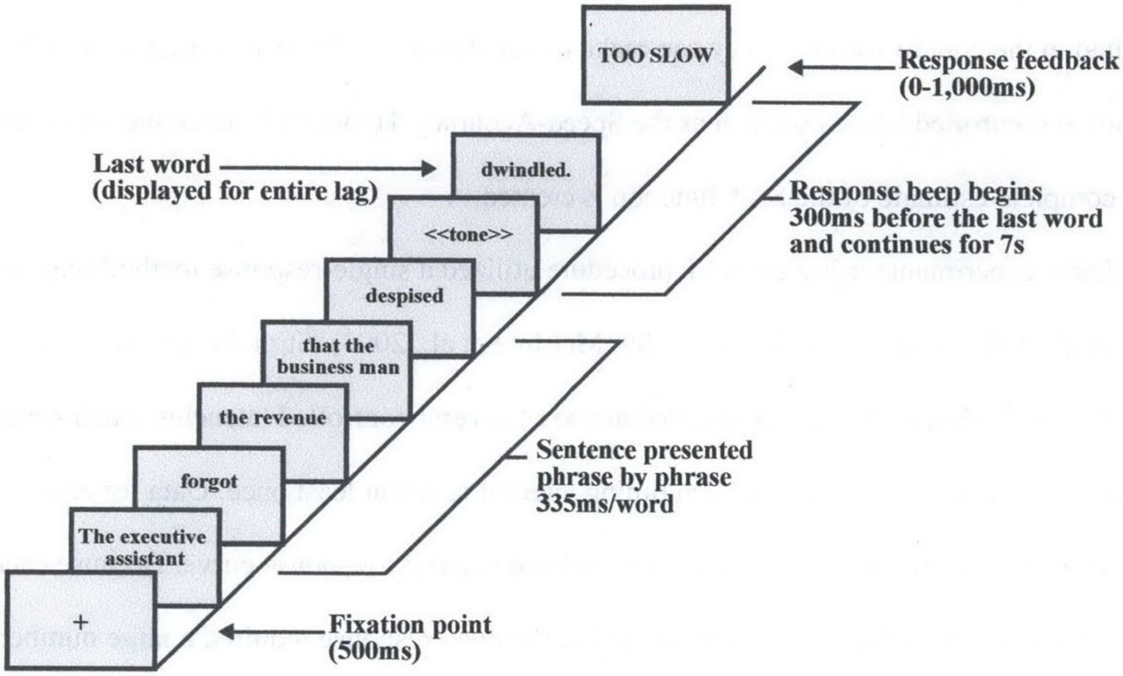


Figure 2: Multiple-Response SAT procedure (adapted from(McElree, 2000)

In the area of language and memory, the SAT procedure has been applied to both list-based and sentence-based tasks. List-based SAT tasks have been used to examine responses to prompts for either item (Was a target in the previously presented list?) or order (When was the target in relation to other items in the list?) information (Dosher, McElree, Hood, & Rosedale, 1989; McElree & Dosher, 1993) . In the sentence-based variant, subjects make plausibility judgments about visually presented sentences. Experimental stimuli are designed so that plausibility cannot be determined before the presentation of the final word. The response period begins before final word presentation, so response accuracy begins at chance. The sentence-based SAT paradigm has been used to evaluate comprehension and retrieval of information such as nouns and modifiers (McElree, Murphy, & Ochoa, 2006) and filler-gap constructions (McElree, Jia, & Litvak, 2000) .

Results from the SAT task are plotted as d-prime (henceforth d') accuracy values (a normalized value of hits minus false alarms) versus time elapsed since presentation of target.

$$d' = \frac{z\ score\ of\ p(yes|yes)}{z\ score\ of\ p(yes|no)}$$

Response curves are created for each subject as well as the entire group. A d' value of 1 or less indicates that responses are not greatly above chance accuracy. Values of d' greater than 1 indicate reasonable above-chance responses. All d' values are calculated automatically in R (R Core Team, 2013).

The resulting data demonstrate a period of time where the subject performs at chance, followed by a steep increase in accuracy, and a gradual tapering to an end accuracy rate.

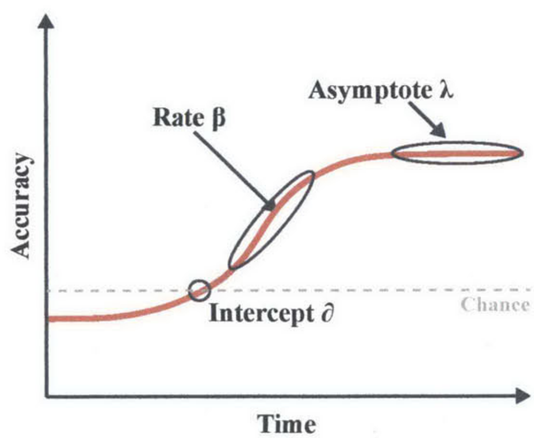


Figure 3: SAT response function

Data are described by three parameters: the intercept, the rate leading to the asymptote, and the asymptote. The intercept is the time point at which d' accuracy crosses 0; that is the point at which the subject performs above chance. The increase in d' value from intercept to asymptote is described by the rate, or the speed at which the subject retrieves information to make an accurate judgment. The asymptote is the plateaued final accuracy with which the subject responds to a stimulus type. Data are fit by an exponential approach to asymptote, where λ refers to asymptote, β refers to rate, and δ refers to intercept. Models are then described in the shortened notation of λ-β-δ, where λ, β, and δ refer to the number of unique values for each parameter:

$$d'(t) = \lambda\left(1 - e^{-\beta(t-\delta)}\right) \, for \; t > \delta, else \; 0$$

Asymptote, or end accuracy, represents the overall accessibility of the relevant piece of information in memory. Both serial and parallel search mechanisms predict that as the number of items in the search set increases, the accessibility of items in memory decreases, which would be reflected in a lower end accuracy. After d' values have been calculated, asymptotic d' values are determined for each sentence type for each subject by determining whether d' changes over the last four time points. If it does not, asymptotic d' is calculated by averaging the final four d' values for each sentence type. Asymptotes are tested for effects of stimulus conditions. Results of this inferential testing determine the number of asymptotes to be used in modeling.

Intercept and rate describe the speed and manner in which an item is accessed. They are the dynamics of the SAT curve, and this is where the differences between serial and parallel search are evident. In parallel search, all accessible items are contacted simultaneously. Therefore, the rate of accuracy increase is equal regardless of item set size. For a serial search mechanism, the rate of retrieval varies based on the size of the item set, resulting in a difference of rate between stimulus types. A difference in the dynamics of the SAT curve, therefore, would suggest the use of a serial search.

Models of performance on stimuli of different types are systematically tested within subjects and for the subject group as a whole, beginning from the null (1λ-1β-1δ) to a fully saturated model (6λ-6β-6δ). Using linear mixed effect modeling (lme4 package in R, Baayen et al. 2014), the goodness of fit of each model to the data is calculated, producing an adjusted R^2

value. The best fitting model is judged to be the one with the largest overall R^2 value, consistent largest R^2 for individuals, and fewest parameters.

A. Subjects

Subjects were native English speakers, aged 18-30, with no reported history of neurological or linguistic deficits, corrected-to-normal vision, and functional hearing. Subjects were recruited through the BU Quickie Jobs service and tested at the Language Science Lab at Sargent College, Boston University. Subjects were compensated at a rate of $10 per hour for participation.

The target usable number of students for this study was 20, consistent with the standard for experiments of this type (Van Dyke & McElree, 2011) .

Of the 54 subjects who began the study, 15 withdrew before completing all eight experimental sessions, and three were lost to follow-up. Initial testing suggested that the task was difficult for subjects to master. Due to the nature of the task and the number of sessions required from participants, some subjects demonstrated inconsistent plausibility judgments. This necessitated the creation of an additional screening measure. Subject data were reviewed following the second experimental session, and if the data did not achieve d' values above 1 for a majority of sentence types or approach an asymptotic accuracy rate, the subject was put on a "watch" list for analysis. If a subject consistently failed to meet these inclusion criteria, they were dropped from the study. 10 subjects were discontinued following this additional screening process. Data from the remaining 25 subjects were analyzed individually for NP/S and NP/Z sentence sets, the same screening procedure was applied for each sentence set, and subjects excluded. This yielded a total n of 17 for the NP/S sentences and 18 for NP/Z sentences. Individual sessions in which subjects fell asleep or reported extreme fatigue (5 half-sessions overall) were discarded.

B. Stimuli

Experimental sentences were designed so that the plausibility of the sentence could not be determined before presentation of the final word. The final word of the sentence requires retrieval of a referent from the beginning of the sentence, and may require a reanalysis due to reattachment of the head NP of the sentence. Each battery followed a 2x3x2 design: ambiguous versus unambiguous, length (short, medium, or long), and plausibility (plausible versus implausible).

In the NP/S sentence type, the main verb may take either a noun phrase (NP) or sentential complement (S) as a modifier. Thus, an ambiguity is induced at the noun phrase following the main verb. The trigger for reanalysis occurs at the final word of the sentence. When the complementizer "that" occurs directly after the main verb, this ambiguity is resolved early: the content after a singular NP must be a sentential complement. Ambiguity, determined by presence or absence of complementizer "that," was a main factor in sentence design for this experiment.

The second main factor in NP/S sentence design was length. Length between the final word of the sentence and its referent was varied through the addition of intervening clauses. In the "short" condition, no clauses are added. The "medium" length sentences contain one additional intervening clauses, and the "long" sentences contain two intervening clauses. All sentence types appeared in both a plausible and implausible variant.

Half the experimental sentences were made implausible at the final word of the sentence or in the main clause. This ensured that subjects paid attention to the entirety of sentences rather than just the beginning and end.

Ambiguous	Short	Plausible	The executive assistant forgot the revenue dwindled.
		Implausible	The executive assistant forgot the revenue fumed.
	Medium	Plausible	The executive assistant forgot the revenue that the businessman despised dwindled.
		Implausible	The executive assistant forgot the revenue that the businessman despised fumed.
		Implausible (main clause)	*The executive assistant forgot the revenue that the businessman angered dwindled.
	Long	Plausible	The executive assistant forgot the revenue that the businessman who made the deal despised dwindled.
		Implausible	The executive assistant forgot the revenue that the businessman who made the deal despised fumed.
		Implausible (main clause)	*The executive assistant forgot the revenue that the businessman who made the deal angered dwindled.
Unambiguous	Short	Plausible	The executive assistant forgot that the revenue dwindled.
		Implausible	The executive assistant forgot that the revenue fumed.
	Medium	Plausible	The executive assistant forgot that the revenue that the businessman despised dwindled.
		Implausible	The executive assistant forgot that the revenue that the businessman despised fumed.
		Implausible (main clause)	*The executive assistant forgot that the revenue that the businessman angered dwindled.
	Long	Plausible	The executive assistant forgot that the revenue that the businessman who made the deal despised dwindled.
		Implausible	The executive assistant forgot that the revenue that the businessman who made the deal despised fumed.
		Implausible (main clause)	*The executive assistant forgot that the revenue that the businessman who made the deal angered dwindled.

Table 1: NP/S stimulus sentence types, controls indicated with asterisk

In the NP/Z construction, ambiguity is created by an initial verb that may take either a noun phrase (NP) or null element (Z). Based on the content following this verb, the original parse will be either adequate or require reanalysis. Reanalysis forces reassignment of the noun that follows the verb as the subject of the main clause. In the early closure sentences, the content following the initial verb must undergo raising from object to subject position. In late closure sentences, initial assignment of the verb is adequate for comprehension. Sturt and colleagues examined this sentence structure with a self-paced reading paradigm and found increased reading times for early closure sentences (Sturt, Pickering, & Crocker, 1999) .

Early Closure	Short	Plausible	Whenever the train left the station emptied.
		Implausible (main clause)	Whenever the train left the station congealed.
	Medium	Plausible	Whenever the train left the station that the tourist visited emptied.
		Implausible (main clause)	Whenever the train left the station that the tourist visited congealed.
		Implausible (relative clause)	*Whenever the train left the station that the tourist cooked emptied.
	Long	Plausible	Whenever the train left the station that the tourist who loved the architecture visited emptied.
		Implausible (main clause)	Whenever the train left the station that the tourist who loved the architecture visited congealed.
		Implausible (relative clause)	*Whenever the train left the station that the tourist who loved the architecture cooked emptied.
Late Closure	Short	Plausible	Whenever the train left the station the platform emptied.
		Implausible (main clause)	Whenever the train left the station the platform congealed.
	Medium	Plausible	Whenever the train left the station that the tourist visited the platform emptied.
		Implausible (main clause)	Whenever the train left the station that the tourist visited the platform congealed.
		Implausible (relative clause)	*Whenever the train left the station that the tourist cooked the platform emptied.
	Long	Plausible	Whenever the train left the station that the tourist who loved the architecture visited the platform emptied.
		Implausible (main clause)	Whenever the train left the station that the tourist who loved the architecture visited the platform congealed.
		Implausible (relative clause)	*Whenever the train left the station that the tourist who loved the architecture cooked the platform emptied.

Table 2: NP/Z stimulus sentence types, control sentences indicated by asterisk

For the NP/Z battery, as in the NP/S, the factor of length between initial verb and final word was manipulated through the addition of clauses between the ambiguous verb and the final word of the sentence. In the "short" condition, no additional clauses were added. "Medium" sentences included an additional subject relative clause, and "long" sentences contained both extra subject and object relative clauses. Each sentence type occurred both a plausible and implausible variant. Experimental sentences were implausible due to errors in the main clause, and control sentences became implausible in the relative clauses.

C. Procedure

Each subject completed a training session lasting approximately one hour in order to become familiarized with the sentence plausibility task and SAT response modality. A brief re-training preceded the start of the first experimental session. Subjects completed eight experimental sessions lasting approximately one hour each, with no more than two sessions allowed in a single day. If more than one week elapsed between experimental sessions, a short re-training session was administered prior to starting the experimental lists.

In each of eight sessions, subjects completed two lists of 72 sentences each, with several short breaks provided during each session. Lists were completed in a randomized order.

During experimental sessions, subjects were seated at a computer in a quiet room. The experiment was run on a personal computer running Windows using E-Prime (Version 2.0; Psychological Software Tools, Pittsburgh, PA). All stimuli were presented on a flat screen monitor as white text on a black background. Subjects performed a key press to begin administration of each sentence. Following presentation of a visual fixation point, sentences were presented on the screen one phrase at a time. Short tones (100 ms in duration) began prior to presentation of final word of sentence. They continued at a rate of one per 350 milliseconds and ended six seconds after presentation of the final word. As responses must begin before plausibility can be determined, subjects were trained to begin all responses with a neutral key press ("yes" and "no" keys simultaneously). They changed to pressing a single key once an acceptability judgment had been made. Subjects were trained to change responses if their judgment changes during the response period.

Visual feedback was provided following each response period if subjects began their responses too late, responded too many or too few times, did not begin their response with a neutral key press, or failed to respond. No feedback regarding accuracy of response was provided.

IV. Results

A. NP/S sentences

17 subjects met criteria for inclusion in NPS analysis. Aggregate data for this sentence type are shown in Figure 4, below. Results for individual subjects may be found in Appendix 1 (NP/S Results by Subject).

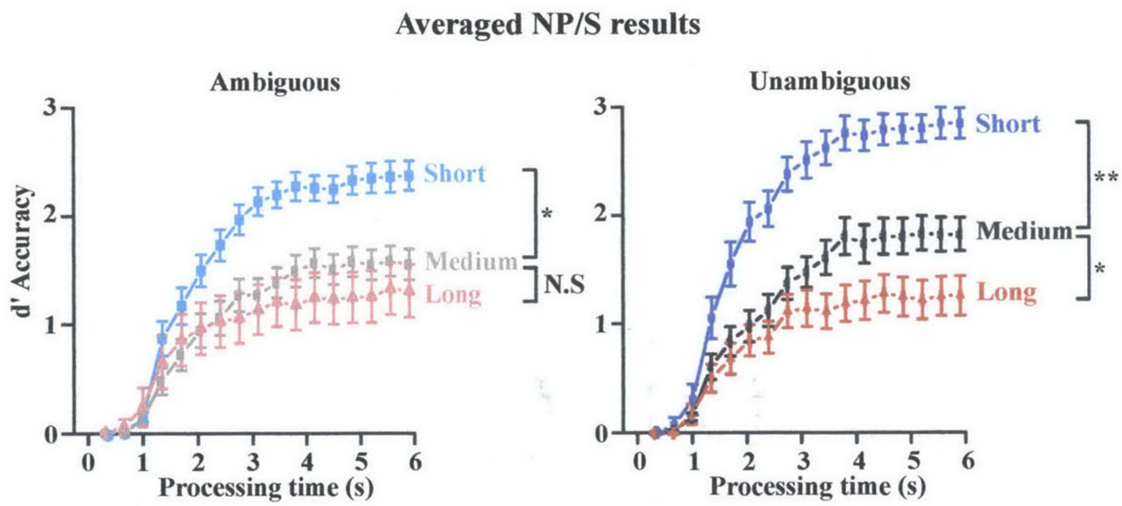


Figure 4: NP/S d' accuracy for all subjects. n=17.
*** significant at $p < 0.05$; ** significant at $p < 0.005$, *** significant at $p < 0.001$**

A two-way ANOVA was used to test for the effect of ambiguity and sentence length on asymptotic d' accuracy. There was a main effect of both ambiguity, $F(1,12) = 6.98$; $p = .021$ and length $F(2,12) = 68.10$; $p < 0.001$. Interaction between ambiguity and sentence length was not significant, $F(2,12) = 2.58$; $p = 0.117$. Pairwise comparisons indicated that asymptotic accuracy for short sentences was significantly greater than medium and long sentences (short vs. medium $p = 0.018$, short vs. long $p = .008$). Differences in asymptotic accuracy of medium and long sentences reached significance in unambiguous sentences, $t(16) = 3.55$, $p = 0.017$, but not in ambiguous sentences, $t(16) = 1.69$, $p = .110$. Asymptotic accuracy for short ambiguous and unambiguous sentences was significantly different, $t(16) = 3.04$, $p = .008$. Accuracy for ambiguous and unambiguous medium-length sentences did not differ significantly, $t(16) = 1.69$, p = .110. Lastly, asymptotic accuracy for ambiguous and unambiguous long sentences did not differ significantly t(16) = 0.16, p = .878. Based on the results of inferential testing of asymptotic values, the primary focus for modeling included models with two (one asymptote assigned to short sentences, one asymptote assigned to medium and long sentence), three (one asymptote assigned to each length), and six asymptotes (one asymptote assigned to each sentence type).

Data were fit by an exponential approach to an asymptote. Table 3 includes adjusted R^2 values for averaged data and summary R^2 information by subject. For further model fit values by subjects, see Appendix 3.

Model	Overall fit	Individual fits M (SE)	Model	Overall fit	Individual fits M (SE)
6λ-6β-6δ	.997	.973 (.003)	**3-1-2 amb**	.969	.891 (.022)
1λ-1β-1δ	.663	.572 (.045)	**3-1-3**	.966	.881 (.023)
2-1-1	.944	.823 (.026)	**3-1-4**	.972	.897 (.021)
2-1-2 SvML	.944	.822 (.037)	**3-1-6**	.972	.903 (.018)
2-1-3	.943	.835 (.033)	**3-2-1 SMvL**	.966	.881 (.023)
2-1-4	.947	.833 (.035)	**3-2-1 SvML**	.965	.883 (.023)
2-1-6	.948	.862 (.022)	**3-2-1 amb**	.980	.915 (.019)
2-2-1 SMvL	.955	.824 (.037)	**3-2-2 SvML**	.981	.881 (.024)
2-2-1 SvML	.969	.863 (.029)	**3-2-2 amb**	.981	.917 (.019)
2-2-2 SvML	.958	.866 (.028)	**3-2-3 SvML**	.981	.882 (.024)
2-2-2 amb	.958	.852 (.034)	**3-2-3 amb**	.981	.916 (.020)
2-2-3 SvML	.957	.870 (.028)	**3-2-6 SvML**	.985	.902 (.019)
2-2-3 amb	.957	.868 (.030)	**3-2-6 amb**	.985	.928 (.016)
2-2-6 SvML	.965	.893 (.026)	**3-3-1**	.967	.897 (.025)
2-2-6 amb	.965	.893 (.020)	**3-3-2 SvML**	.970	.894 (.024)
2-3-1	.955	.869 (.026)	**3-3-2 amb**	.970	.894 (.023)
2-3-2 SvML	.958	.891 (.025)	**3-3-3**	.967	.883 (.024)
2-3-2 amb	.958	.877 (.025)	**3-3-6**	.972	.907 (.018)
2-3-3	.960	.874 (.025)	**3-4-1**	.987	.937 (.010)
2-3-6	.965	.896 (.018)	**3-4-3**	.987	.942 (.009)
2-4-1	.961	.880 (.023)	**3-6-1**	.988	.955 (.007)
2-4-2	.897	.883 (.022)	**3-6-2 SvML**	.990	.954 (.007)
2-4-3	.891	.895 (.017)	**3-6-2 amb**	.990	.952 (.007)
2-4-6	.894	.909 (.013)	**3-6-3**	.988	.906 (.044)
2-6-1	.974	.930 (.013)	**3-6-6**	.989	.967 (.016)
2-6-2 SvML	.976	.932 (.013)	**6-1-1**	.994	.961 (.006)
2-6-2 amb	.976	.933 (.012)	**6-2-1 SMvL**	.995	.962 (.006)
2-6-3	.980	.940 (.009)	**6-2-1 SvML**	.994	.964 (.006)

2-6-6	.981	.938 (.013)	**6-2-1 amb**	.994	.963 (.005)
3-1-1	.965	.881 (.023)	**6-3-1**	.996	.966 (.006)
3-1-2 SvML	.969	.882 (.023)	**6-6-1**	.997	.970 (.004)

Table 3: Selected NP/S model fits (values in adjusted R^2). All 3-X-X models are assigned asymptotes based on length (one asymptote for short sentences, one asymptote for medium sentences, and one asymptote for long sentences). 2-X-X models assign one asymptote to short sentences and a second shared asymptote to medium and long sentences. In 6-X-X models, each sentence type is assigned a unique parameter value. In models with 1 value for a parameter, all sentence types share a single parameter value. In X-2-X and X-X-2 models, parameter values assigned by either ambiguity or sentence length. In models with names ending in "amb," one value is assigned to ambiguous sentences, and the other to ambiguous sentences. In other cases, rate or intercept are assigned based on length. Models ending in SvML group together medium and long length sentences. Models ending in SMvL group together short and medium length sentences. In X-4-X and X-X-4 models, unambiguous sentences are grouped together, and each ambiguous sentence type is assigned a unique parameter value, for a total of 4 unique values.

To provide basis for comparison, both the null (1λ-1β-1δ) and fully saturated (6λ-6β-6δ) models were run for each subject as well as aggregate data. For the 1λ-1β-1δ model, adjusted R^2 values for individuals ranged from .190-.850, with an average of .572. In the fully saturated model, individual R^2 values ranged from .944-.989, with an average of .973.

The models with the highest adjusted R^2 values for both individual and aggregate data were the 6λ-6β-6δ and 6λ-6β-1δ models. Individual fits for the 6λ-6β-1δ model ranged from .922-.988, with an average of .970. When considering the top three R^2 values for all subjects, nine of 17 subjects were best fit by the 6λ-6β-6δ model, and nine subjects were fit best by the 6λ-6β-1δ model. Two additional subjects were fit best by three-asymptote models. Both 6λ-6β-6δ and 6λ-6β-1δ models produced adjusted R^2 values significantly higher than the null 1λ-1β-1δ (paired t-test, $p < .001$, $p < .001$). Reducing the number of intercept parameters did not yield a significant difference in adjusted R^2 (6λ-6β-6δ versus 6λ-6β-1δ, paired t-test, $p = .540$). Reducing the number of both rate and intercept parameters also did not yield a significant decrease in adjusted R^2 (6λ-6β-6δ versus 6λ-1β-1δ, paired t-test, $p = .091$). Reducing the number of asymptotes resulted in poorer adjusted R^2 values (6λ-1β-1δ versus 3λ-1β-1δ, $p = .001$).

Each model produces a set of estimated values for each parameter (asymptote, rate, intercept) for each subject as well as aggregate data. These estimated values produce a secondary means for comparison of goodness of fit for each model. A one-way ANOVA was used to test for estimated rate value differences between the six sentence types in the 6λ-6β-1δ model. Estimated rates did not differ significantly across sentence types, $F(96,5) = 2.093$, $p = 0.073$, further supporting the adequacy of a model with a single rate parameter. Additionally, estimated values for the rate parameter did not differ significantly between 6λ-6β-1δ and 6λ-1β-1δ ($p = .075$). Although adjusted R^2 values were significantly reduced with reduction in number of asymptotes (6λ-1β-1δ versus 3λ-1β-1δ), estimated asymptotic values for these two models were not significantly different (paired t-test, $p = .896$). This, combined with inferential testing of observed asymptotic d' values, indicates that fewer asymptotes may not harm the fit of the model.

Based on individual and aggregate R^2 values, the 6λ-1β-1δ model was identified as the best-fit and least-saturated model for the data. Some support for assignment of a model with fewer asymptotes emerged in analysis of estimated parameter values, but the 6λ-1β-1δ model was ultimately selected for its high R^2 value.

B. NP/Z sentences

18 subjects met criteria for inclusion in NP/Z analysis. Aggregate results for this sentence type are shown in Figure 5. Results for individual subjects may be found in Appendix 2.

A two-way ANOVA was used to test for the effect of closure type and sentence length on asymptotic d'. There a main effect of closure type, $F(1,13) = 63.57$; $p < .001$, and length, $F(2,13) = 49.44$; $p < .001$. Interaction between the effects of ambiguity and sentence length was significant, $F(2,13) = 28.48$; $p < .001$.

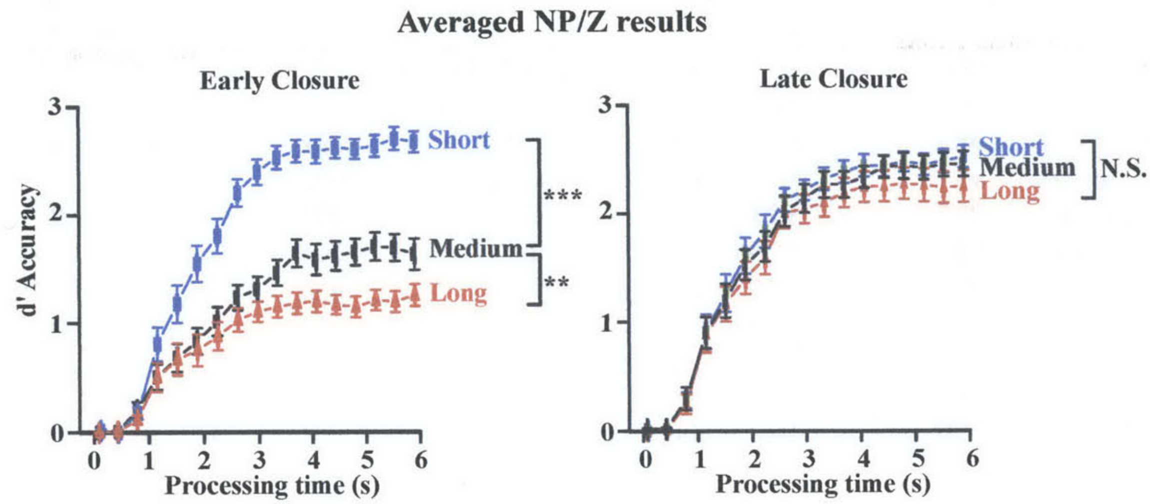


Figure 5: NP/Z d' accuracy for all subjects. n=18.
*** significant at p < 0.05; ** significant at p < 0.005, *** significant at p < 0.001**

Pairwise comparisons indicated that asymptotic accuracy for early closure short sentences wassignificantly higher than early closure medium-length sentences, $t(17) = 8.33$, $p < .001$, which in turn had a significantly higher asymptotic d' accuracy than early closure long sentences $t(17) = 3.88$, $p = 0.001$. All differences within the late closure sentences were insignificant (late closure short versus medium, $t(17) = 0.37$, $p = .716$, short versus long $t(17) = 1.87$, $p = .078$, medium versus long $t(17) = 1.50$, $p = .151$). Asymptotic accuracy for early and late closure short sentences did not differ significantly, $t(17) = 1.41$, $p = .177$. Asymptotic d' accuracy for early and late closure medium-length sentences did differ significantly, $t(17) = 6.44$, $p < .001$. Finally, asymptotic accuracy for early and late long sentences differed significantly, $t(17) = 8.78$, $p < .001$. Based on these results, modeling focused on models with three asymptotes (one shared asymptote assigned to short early closure sentences and all late closure sentences, one unique asymptote for medium early closure sentences, and one unique asymptote for long early closure sentences). Additional models with four asymptotes (unique asymptotes assigned to each early closure sentence, one asymptote assigned to all late closure sentences) and six asymptotes (one unique asymptote assigned to each sentence type and length) were included for comparison.

Model	Overall fit	Individual fits M (SE)	Model	Overall fit	Individual fits M (SE)
6λ-6β-6δ	.996	.972 (.003)	**3-6-2 closure**	.991	.952 (.007)
1-1-1	.769	.702 (.028)	**3-6-3 asym**	.990	.949 (.007)
3-1-1	.985	.922 (.010)	**3-6-3 length**	.991	.952 (.007)
3-1-2 closure	.985	.923 (.010)	**3-6-4**	.992	.952 (.007)
3-1-3 asym	.985	.923 (.010)	**3-6-6**	.992	.954 (.007)
3-1-3 length	.985	.926 (.009)	**4-1-1**	.991	.940 (.009)
3-1-4	.985	.926 (.010)	**4-1-2**	.992	.941 (.009)
3-1-6	.985	.930 (.009)	**4-1-3 asym**	.991	.942 (.009)
3-2-1 closure	.986	.930 (.010)	**4-1-4**	.992	.943 (.009)
3-2-1 SMvL	.987	.933 (.008)	**4-2-1**	.992	.942 (.009)
3-2-1 SvML	.989	.932 (.008)	**4-3-1 asym**	.992	.944 (.009)
3-2-2 closure	.987	.933 (.010)	**4-4-1**	.992	.945 (.009)
3-2-3 closure	.872	.932 (.010)	**4-4-2 asym**	.992	.946 (.009)
3-2-4 closure	.988	.935 (.010)	**4-4-3 asym**	.992	.945 (.009)
3-2-6 closure	.988	.940 (.008)	**4-4-4**	.992	.946 (.009)
3-3-1 asym	.986	.925 (.010)	**4-4-6**	.993	.950 (.008)
3-3-1 length	.989	.940 (.007)	**4-6-1**	.994	.960 (.005)
3-3-2 asym	.986	.927 (.010)	**4-6-6**	.995	.962 (.005)
3-3-3 asym	.985	.925 (.010)	**6-1-1**	.994	.963 (.003)
3-3-3 length	.986	.931 (.009)	**6-1-2 closure**	.995	.965 (.003)
3-3-4 asym	.985	.927 (.010)	**6-1-3 asym**	.994	.965 (.003)
3-3-6 asym	.986	.932 (.009)	**6-1-4**	.995	.968 (.003)
3-4-1	.988	.934 (.010)	**6-1-6**	.995	.968 (.003)
3-4-2 closure	.989	.937 (.010)	**6-2-1 SMvL**	.994	.964 (.004)
3-4-3 asym	.988	.934 (.010)	**6-2-1 SvML**	.994	.965 (.003)
3-4-3 length	.988	.938 (.009)	**6-3-1 length**	.995	.966 (.003)
3-4-4	.989	.936 (.010)	**6-4-1**	.996	.969 (.003)
3-4-6	.990	.941 (.009)	**6-6-1**	.996	.972 (.003)
3-6-1	.990	.949 (.007)			

Table 4: Selected NP/Z model fits (all values in adjusted R^2). All 3-X-X models are assigned a shared asymptote for early closure short and all late closure sentences. Early closure medium and long sentences are assigned unique asymptotes. For X-3-X and X-X-3 models, if model name ends in "length," rate or intercept are assigned according to sentence length. Model names ending in "asym" or without additional labeling denote models where rate or intercept are assigned in the same manner as 3-asymptote models (shared value for early closure short and all late closure, unique values for early closure medium and early closure long). In models with 4 asymptotes, rates, or intercepts, each length of early closure sentences is assigned a unique value, and all late closure sentences share a single set of parameter values. Models with 2 rates or intercepts have parameter values either assigned by closure (model name ends in "closure") or length (model name ends in "SvML" or "SMvL").

As with NP/S sentences, data were fit to an exponential approach to asymptote model. Table 4 includes adjusted R^2 values for aggregate data as well as summary R^2 data for individual subjects. For further results of modeling by subject, see Appendix 4.

Overall R^2 fit for the null 1λ-1β-1δ model was .769. Individual fits ranged from .481 to .911 with an average R^2 of .702. Overall fit for the saturated 6λ-6β-6δ model was .996, with individual fits ranging from adjusted R^2 values of .939 to .988, and averaging 0.973.

The models with the highest adjusted R^2 values for the most subjects were the fully saturated 6λ-6β-6δ model, the 6λ-6β-1δ model, and the 6λ-4β-1δ model. In the 6λ-4β-1δ model, each early closure sentence was assigned a unique rate, and all late closure sentences were assigned to a single rate. When considering the top three R^2 values for all subjects, 16 of 18 subjects were fit best by the 6λ-6β-6δ, 15 by the 6λ-6β-1δ model, and seven by the 6λ-4β-1δ model. The 6λ-6β-6δ model, 6λ-6β-1δ model, and 6λ-4β-1δ models all yielded adjusted R^2 values significantly higher than the null (paired t-test, $p < .001$, $p < .001$, $p < .001$). Reducing the number of intercepts did not significantly decrease adjusted R^2 (6λ-6β-6δ versus 6λ-6β-1δ, paired t-test $p = .660$). Reducing the number of rate parameters from six significantly reduced adjusted R^2 (6λ-6β-1δ versus 6λ-4β-1δ, paired t-test $p = .004$). Reduction in the number of asymptotes significantly changed adjusted R^2 (6λ-6β-1δ versus 4λ-6β-1δ, paired t-test, $p < .001$).

Analysis of estimated parameter values supports some observations of patterns in R^2. Although models with fewer than six rates yielded adjusted R^2 values significantly lower than those with six rates, estimated rate values did not differ significantly when the number of rate parameters was reduced (6λ-6β-1δ versus 6λ-4β-1δ, paired t-test, $p = .101$, 6λ-6β-1δ versus 6λ-3β-1δ, $p = .082$). Estimate rate values were significantly different when the model was reduced to a single rate (6λ-6β-1δ versus 6λ-1β-1δ, $p = .001$). A one-way ANOVA was used to test for estimated rate value differences between the six sentence types in the 6λ-6β-1δ model. Estimated rates did not differ significantly across sentence types, $F(102,5) = 0.517$, $p = .763$. Similar results were yielded by a one-way ANOVA for estimated rate values in the 6λ-4β-1δ model, F(68,3)= 0.578, $p = .632$. Estimated asymptotic values were not significantly different between the 6λ-6β-1δ model the 4λ-6β-1δ model, paired t-test, $p = .146$, but they were between the 6λ-6β-1δ and 3λ-6β-1δ models, $p = .003$).

Based on individual and aggregate adjusted R^2 values, the 6λ-6β-1δ model was identified as the best-fit and least saturated model for the data. Some support for assignment of a model with fewer asymptotes and rates emerged in analysis of estimated parameter values, but the 6λ-6β-1δ model was ultimately selected for its high R^2 value.

V. Discussion

In these experiments, a Speed-Accuracy Tradeoff paradigm (Wickelgren, 1977) was used to explore patterns of information retrieval for sentence comprehension. Sentences varied in ambiguity, syntactic closure type, and length (number of embedded clauses). Subjects made acceptability judgments for all sentences, and data were collected for the time course of response accuracy.

Results are compared with predictions made by the Caplan and Waters bipartite model of sentence comprehension (2013) . According to this model, incoming linguistic information is assigned meaning by a skilled parser. If this parse fails, output of the controlled processor is used instead. The skilled parser is associated with a content-addressable, parallel search mechanism; the controlled process with serial search. The bipartite model predicts that if a sentence is adequately complex or long to cause failure by the parser, response to comprehension tasks will show evidence of serial search by the parser; namely, differences in the dynamics of information retrieval.

Consistent with the results of previous work, asymptotic accuracy (the accuracy of accessing information given adequate time), decreased both with increasing sentence length and increased syntactic ambiguity. In the NP/S construction, the shortest sentences (with no intervening clauses between NP and referent) had a significantly higher asymptotic accuracy rate than sentences with greater distance between the two items. Asymptotic accuracy differed between medium and long sentences only in the unambiguous case. In the case of the NP/Z construction, length had a greater effect on asymptotic accuracy in sentences with early closure constructions, and thus, greater syntactic ambiguity. All late closure sentences, regardless of length, had equivalent asymptotes. The shortest of the early closure sentences patterned together with late closure sentences. Early closure sentences of medium length had lower end accuracy rates than the short sentences, and long early closure sentences an even lower end accuracy rate. This suggests that closure type only becomes a significant factor in comprehension accuracy after a critical point of sentence length (in this case, the medium-length sentence).

Use of a serial search mechanism predicts a difference in retrieval dynamics for sentences of varying length and syntactic complexity: in the NP/S sentence type, this pattern was not found to occur. While there is evidence that a model with fewer than six asymptotes may best fit the data, both R^2 and estimate rate constants indicate that a model with a single rate and single intercept is the best fit. Therefore, there is no evidence for serial search in the NP/S construction sentences.

Results of the NP/Z sentences are ambiguous as to whether or not they support the use of a serial search mechanism. As with NP/S sentences, multiple asymptote parameters are required, and a single intercept yields the highest R^2 value. Models with multiple rates, unlike the NP/S sentences, yield significantly higher R^2 values than those with a single rate. Despite the R^2 advantage of multiple-rate models for the NP/Z construction, estimated rate values were not significantly different for sentence lengths and closure conditions in these multiple-rate models, and these rate values did not differ significantly from those produced by models with a single rate parameter for all sentences.

It is worth nothing that maximal R^2 values were achieved when multiple rates were assigned within the early closure condition. Early closure NP/Z sentences require a more substantial syntactic reanalysis than the late closure NP/Z sentences. If the parser/processor model is correct, it would predict that serial search would be more likely in these early closure sentences. Indeed, assignment of unique rates to sentences within this condition improved overall model fit.

Based solely on R^2 values, the 6λ-6β-1δ and 6λ-4β-1δ models fit NP/Z data best. If the 6λ-4β-1δ model were identified as the best fit for the data, this would indicate that the rate of information retrieval varies when significant syntactic reanalysis is required (a serial search interpretation of the data).

The NP/Z data could also be used to support a parallel, content-addressable search analysis. Although multiple rate parameters yielded high R^2 values, estimated parameter values did not vary significantly between models with one and multiple rate values. Furthermore, estimated rates in the multiple-rate models do not vary systematically with increased length and syntactic complexity/ambiguity. This evidence supports the use of a single rate, and thus the parallel-search model.

All in all, the results of this study indicate that sentence length and syntactic ambiguity have an effect on sentence comprehension accuracy, that retrieval in NP/S sentences is consistent with a parallel search mechanism, and that comprehension of sentences requiring intense syntactic reanalysis (early closure NP/Z sentences) demonstrate retrieval dynamics consistent with either serial or parallel search. Future work may elucidate the exact nature of retrieval mechanisms in this NP/Z construction.

VI. References

Caplan, D., & Waters, G. (2013). Memory mechanisms supporting syntactic comprehension. *Psychonomic Bulletin & Review*, , 1-26.

Dosher, B. A., McElree, B., Hood, R. M., & Rosedale, G. (1989). Retrieval dynamics of priming in recognition memory: Bias and discrimination analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 868.

Henson, R. N. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36*(2), 73-137.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10*(10), 447-454.

Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(5), 1231.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29*(2), 111-123.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General, 118*(4), 346.

McElree, B., & Dosher, B. A. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General, 122*(3), 291.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*(1), 67-91.

McElree, B., Jia, G., & Litvak, A. (2000). The time course of conceptual processing in three bilingual populations. *Journal of Memory and Language, 42*(2), 229-254.

McElree, B., Murphy, G. L., & Ochoa, T. (2006). Time course of retrieving conceptual information: A speed-accuracy trade-off study. *Psychonomic Bulletin & Review, 13*(5), 848-853.

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rinberg, D., Koulakov, A., & Gelperin, A. (2006). Speed-accuracy tradeoff in olfaction. *Neuron, 51*(3), 351-358.

Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language, 40*(1), 136-150.

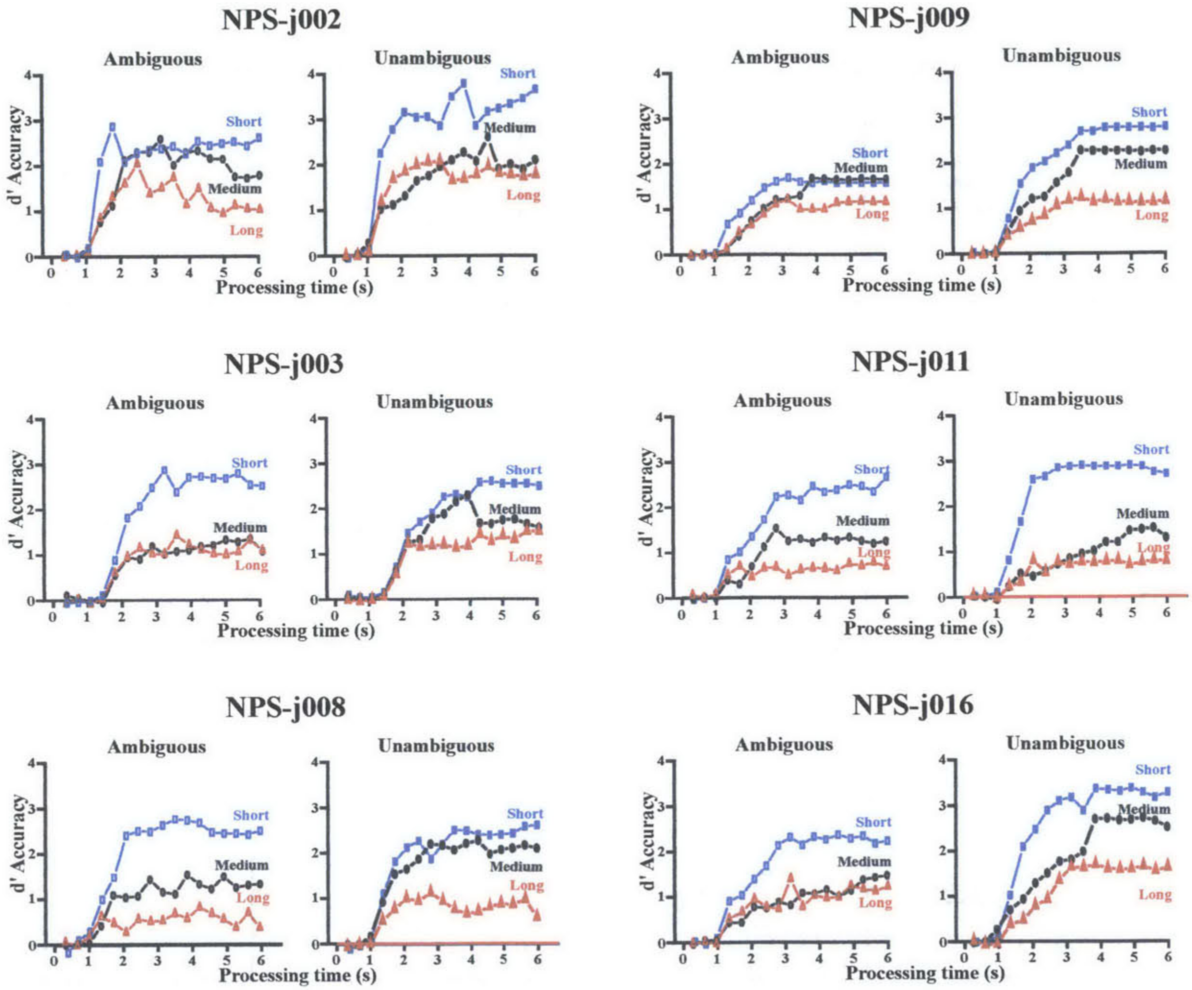
Sutter, A., & Graham, N. (1995). Investigating simple and complex mechanisms in texture segregation using the speed-accuracy tradeoff method. *Vision Research, 35*(20), 2825-2843.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language, 65*(3), 247-263.

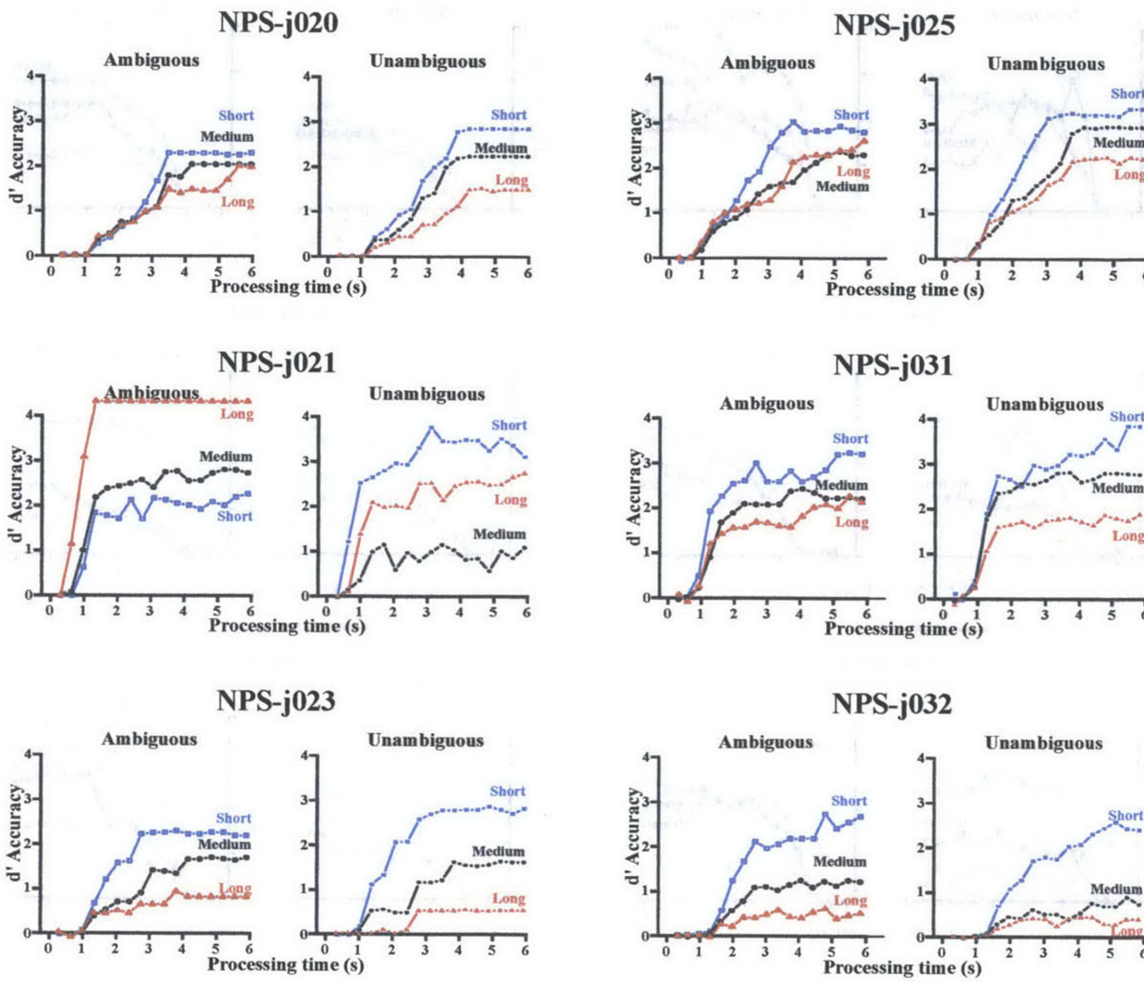
Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*(1), 67-85.

VII. Appendix 1: Individual NP/S results

NPS All Subjects

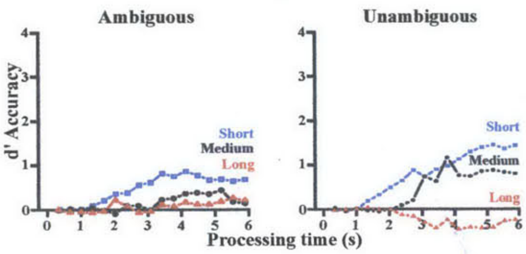


NPS All Subjects

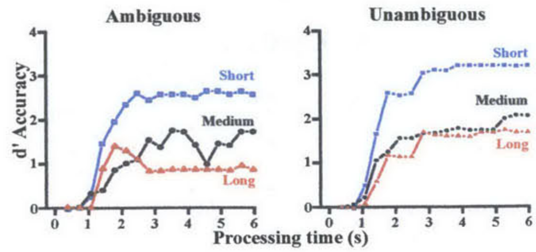


NPS All Subjects

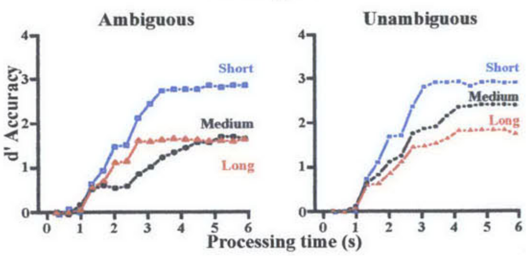
NPS-j035



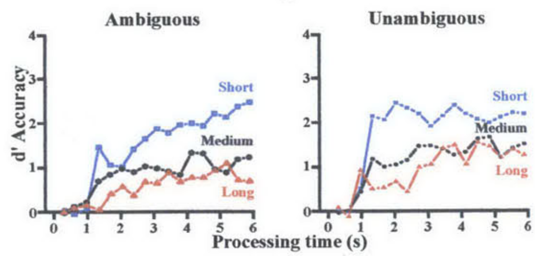
NPS-j049



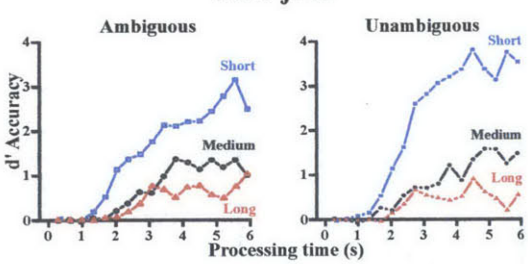
NPS-j038



NPS-j054



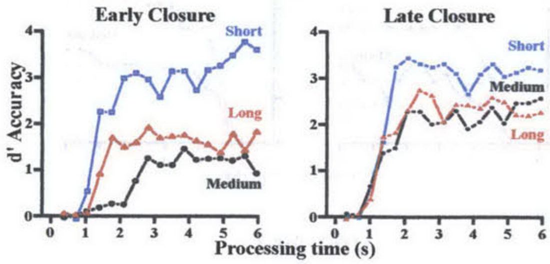
NPS-j043



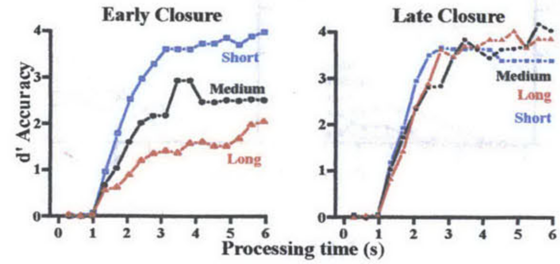
VIII. Appendix 2: Individual NP/Z results

NPZ All Subjects

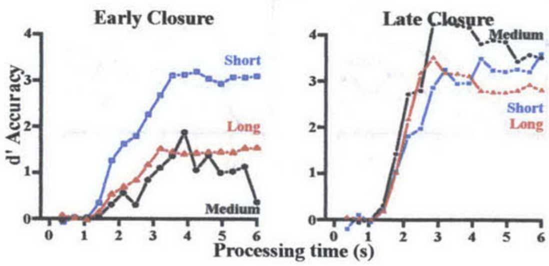
NPZ-j002



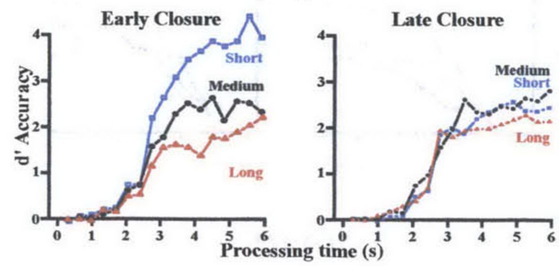
NPZ-j009



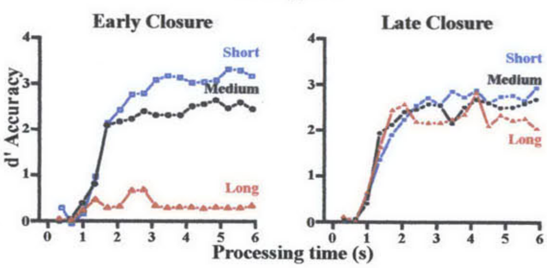
NPZ-j003



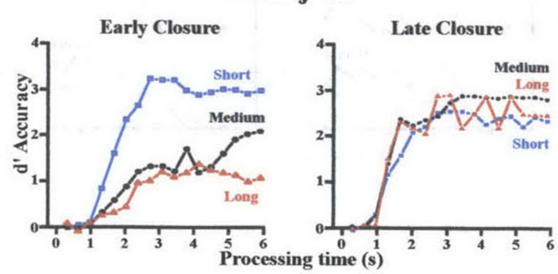
NPZ-j010



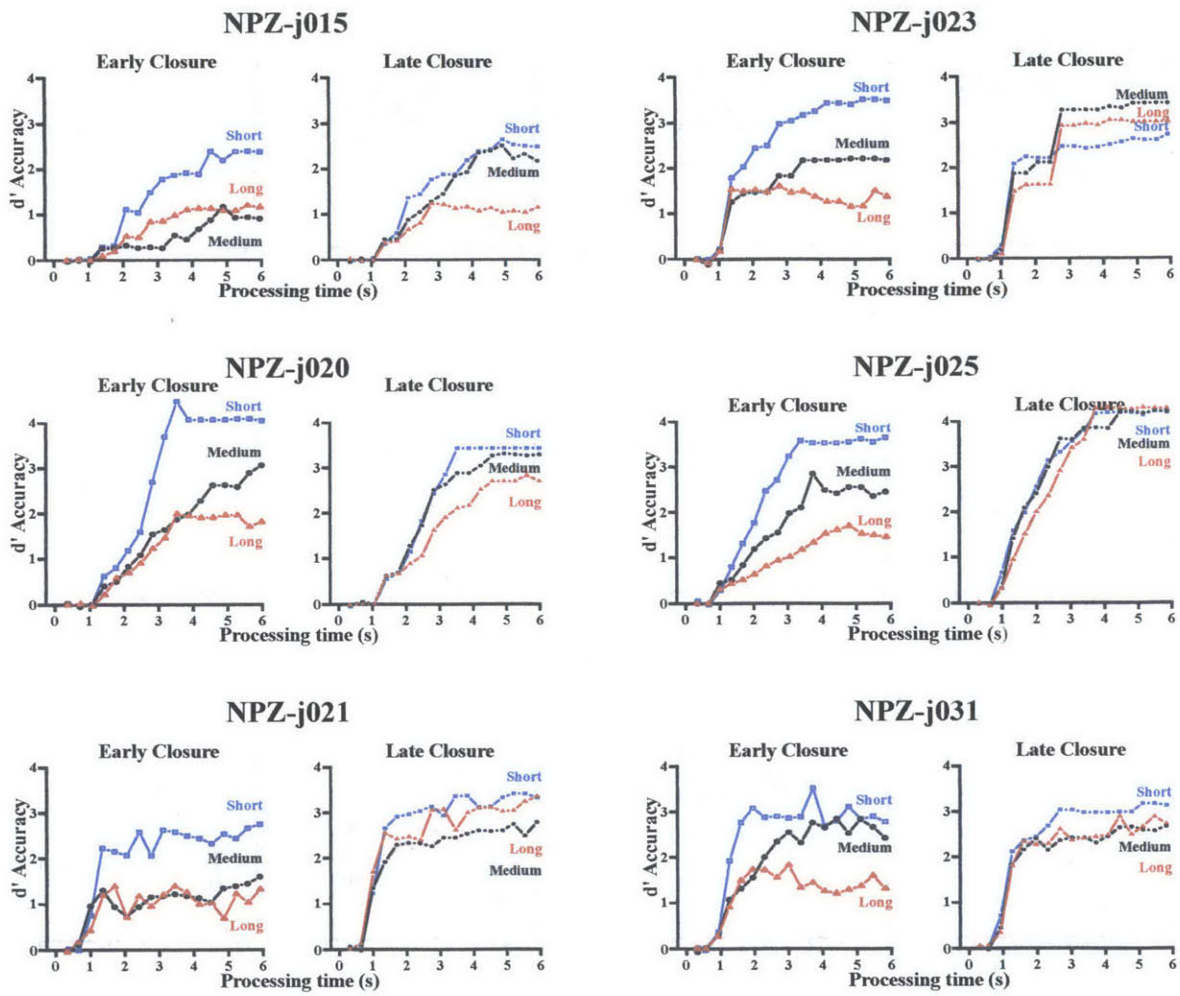
NPZ-j008



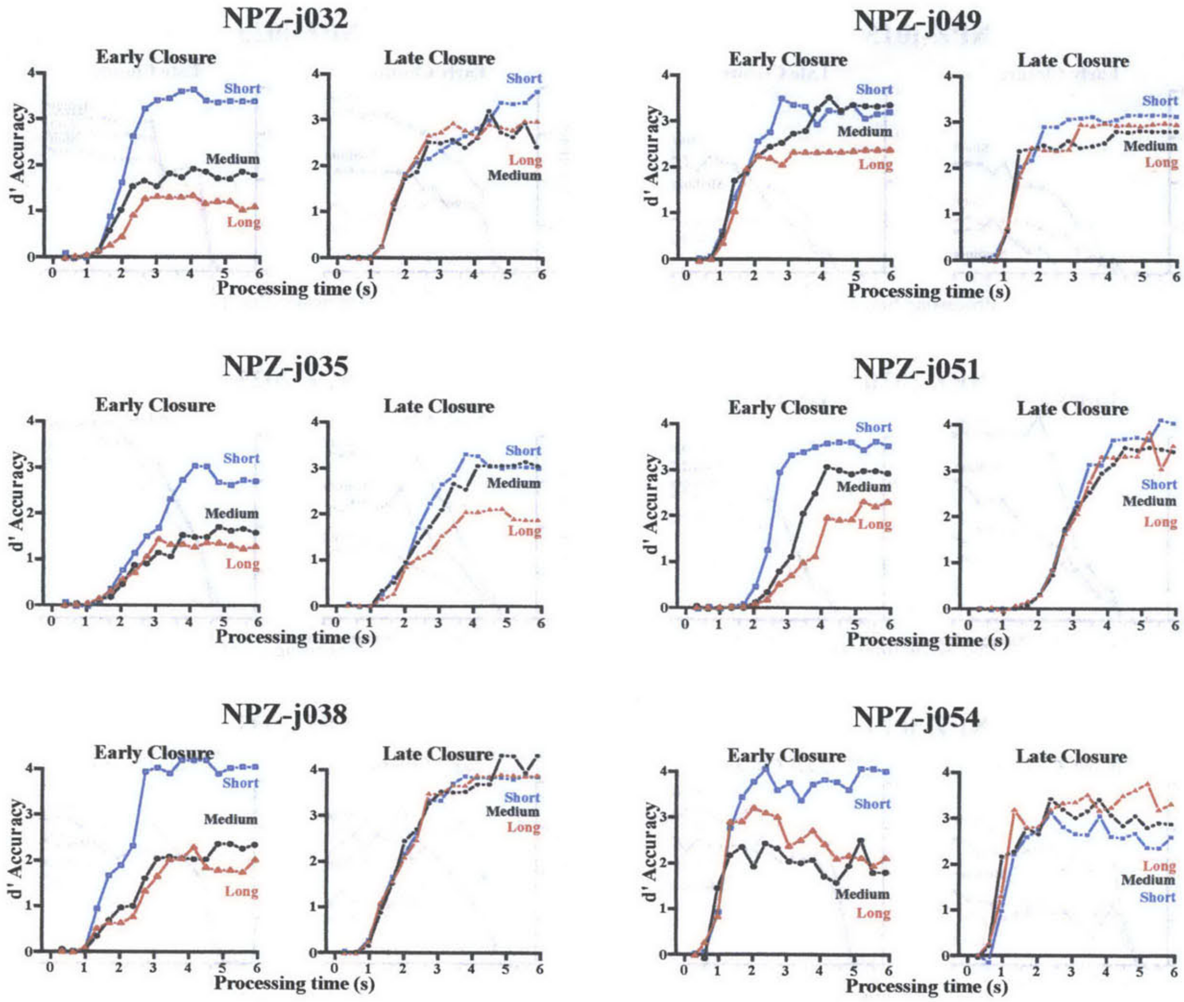
NPZ-j011



NPZ All Subjects



NPZ All Subjects



IX. Appendix 3: Adjusted R^2 data for NP/S fits

Subj. no	666	111	311	312 asym	312 amb	313	314	316	321 SMvL
s2	0.945	0.638	0.870	0.874	0.869	0.876	0.871	0.873	0.869
s3	0.976	0.688	0.936	0.938	0.936	0.938	0.936	0.941	0.939
s8	0.975	0.444	0.916	0.916	0.918	0.916	0.916	0.916	0.918
s9	0.944	0.638	0.849	0.853	0.869	0.853	0.873	0.874	0.848
s11	0.976	0.688	0.943	0.943	0.946	0.944	0.945	0.953	0.943
s16	0.975	0.444	0.805	0.805	0.834	0.803	0.850	0.850	0.803
s20	0.989	0.670	0.940	0.939	0.943	0.938	0.949	0.952	0.941
s21	0.985	0.396	0.590	0.586	0.588	0.583	0.606	0.663	0.588
s23	0.980	0.557	0.942	0.947	0.944	0.946	0.946	0.957	0.942
s25	0.965	0.823	0.934	0.934	0.944	0.934	0.947	0.947	0.934
s31	0.975	0.365	0.941	0.940	0.941	0.940	0.943	0.942	0.940
s32	0.985	0.444	0.957	0.957	0.960	0.957	0.961	0.962	0.958
s35	0.976	0.850	0.749	0.761	0.813	0.758	0.848	0.850	0.746
s38	0.967	0.751	0.937	0.936	0.941	0.937	0.959	0.958	0.937
s43	0.985	0.373	0.925	0.927	0.937	0.926	0.936	0.937	0.924
s49	0.964	0.190	0.911	0.911	0.912	0.910	0.915	0.914	0.912
s54	0.985	0.761	0.825	0.823	0.852	0.822	0.858	0.855	0.835
ALL	0.997	0.663	0.965	0.969	0.969	0.966	0.972	0.972	0.966

Subj. no	321 SvML	321 amb	322 asym	322 amb	323 asym	323 amb	326 asym	326 amb	331
s2	0.875	0.871	0.869	0.870	0.874	0.875	0.872	0.872	0.947
s3	0.939	0.936	0.938	0.936	0.938	0.938	0.941	0.941	0.978
s8	0.916	0.920	0.917	0.919	0.916	0.919	0.920	0.923	0.977
s9	0.854	0.935	0.846	0.941	0.852	0.938	0.874	0.946	0.881
s11	0.944	0.957	0.942	0.959	0.944	0.958	0.953	0.969	0.939
s16	0.810	0.933	0.801	0.941	0.801	0.936	0.849	0.951	0.917
s20	0.940	0.946	0.943	0.947	0.942	0.945	0.954	0.958	0.855
s21	0.586	0.623	0.583	0.638	0.576	0.621	0.634	0.702	0.947
s23	0.955	0.948	0.942	0.948	0.947	0.951	0.956	0.964	0.809
s25	0.940	0.957	0.936	0.957	0.937	0.959	0.948	0.960	0.940
s31	0.940	0.947	0.940	0.949	0.939	0.946	0.941	0.949	0.584
s32	0.958	0.969	0.957	0.970	0.957	0.969	0.962	0.972	0.954
s35	0.750	0.875	0.746	0.880	0.758	0.888	0.849	0.917	0.939
s38	0.937	0.952	0.937	0.952	0.936	0.952	0.958	0.964	0.940
s43	0.926	0.961	0.925	0.964	0.926	0.962	0.939	0.966	0.958
s49	0.911	0.923	0.913	0.928	0.912	0.922	0.915	0.926	0.747
s54	0.826	0.896	0.836	0.897	0.834	0.894	0.866	0.893	0.940
ALL	0.965	0.980	0.981	0.981	0.981	0.981	0.985	0.985	0.967

Subj. no	332 asym	332 amb	333	336	341	343	361	362 asym	362 amb
s2	0.901	0.880	0.880	0.876	0.869	0.873	0.945	0.898	0.881
s3	0.971	0.939	0.938	0.941	0.956	0.961	0.978	0.977	0.964
s8	0.975	0.918	0.915	0.919	0.953	0.955	0.976	0.975	0.955
s9	0.880	0.874	0.853	0.873	0.953	0.956	0.882	0.881	0.959
s11	0.939	0.949	0.946	0.954	0.964	0.972	0.964	0.964	0.980
s16	0.916	0.846	0.806	0.871	0.948	0.950	0.955	0.955	0.967
s20	0.854	0.943	0.942	0.953	0.956	0.956	0.955	0.956	0.963
s21	0.946	0.582	0.572	0.666	0.818	0.852	0.977	0.977	0.928
s23	0.808	0.956	0.954	0.963	0.953	0.960	0.959	0.959	0.974
s25	0.939	0.950	0.941	0.957	0.965	0.965	0.959	0.959	0.973
s31	0.577	0.940	0.939	0.942	0.950	0.949	0.921	0.923	0.952
s32	0.954	0.960	0.957	0.965	0.975	0.976	0.973	0.973	0.984
s35	0.941	0.810	0.761	0.848	0.893	0.907	0.972	0.975	0.896
s38	0.940	0.946	0.940	0.964	0.982	0.981	0.950	0.950	0.983
s43	0.957	0.937	0.926	0.938	0.967	0.970	0.984	0.984	0.972
s49	0.762	0.914	0.913	0.916	0.928	0.926	0.894	0.924	0.933
s54	0.940	0.860	0.834	0.865	0.902	0.901	0.983	0.984	0.917
ALL	0.970	0.970	0.967	0.972	0.987	0.987	0.988	0.990	0.990

Subj. no	363	366	211	212	213	214	216	221 SMvL	221 SvML
s2	0.880	0.883	0.849	0.852	0.851	0.849	0.847	0.853	0.848
s3	0.208	0.722	0.931	0.933	0.932	0.929	0.934	0.933	0.930
s8	0.950	0.897	0.771	0.770	0.772	0.792	0.795	0.770	0.892
s9	0.880	0.883	0.774	0.776	0.780	0.787	0.795	0.776	0.832
s11	0.964	0.968	0.913	0.913	0.912	0.917	0.918	0.914	0.934
s16	0.955	0.961	0.783	0.783	0.784	0.807	0.815	0.787	0.801
s20	0.956	0.966	0.905	0.904	0.921	0.911	0.935	0.904	0.940
s21	0.976	0.980	0.360	0.353	0.379	0.380	0.597	0.353	0.475
s23	0.959	0.968	0.852	0.855	0.887	0.851	0.900	0.861	0.940
s25	0.962	0.963	0.930	0.930	0.929	0.939	0.940	0.935	0.935
s31	0.923	0.931	0.889	0.888	0.888	0.887	0.888	0.888	0.913
s32	0.973	0.973	0.921	0.921	0.927	0.928	0.931	0.921	0.952
s35	0.975	0.974	0.572	0.574	0.696	0.586	0.765	0.571	0.698
s38	0.950	0.951	0.935	0.935	0.934	0.946	0.949	0.935	0.935
s43	0.983	0.985	0.897	0.898	0.904	0.909	0.918	0.897	0.922
s49	0.923	0.935	0.897	0.896	0.896	0.896	0.896	0.896	0.897
s54	0.984	0.984	0.805	0.803	0.815	0.847	0.838	0.807	0.836
ALL	0.988	0.989	0.944	0.944	0.943	0.947	0.948	0.955	0.969

Subj. no	222 asym	222 amb	223 asym	223 amb	226 asym	226 amb	231	232 asym	232 amb
s2	0.851	0.848	0.849	0.851	0.846	0.846	0.851	0.850	0.850
s3	0.932	0.930	0.932	0.931	0.934	0.934	0.933	0.932	0.933
s8	0.891	0.771	0.901	0.774	0.905	0.905	0.891	0.890	0.892
s9	0.832	0.852	0.841	0.857	0.861	0.861	0.836	0.835	0.855
s11	0.933	0.931	0.940	0.928	0.948	0.948	0.936	0.936	0.939
s16	0.799	0.908	0.799	0.909	0.846	0.846	0.803	0.803	0.841
s20	0.939	0.909	0.942	0.927	0.953	0.954	0.939	0.939	0.940
s21	0.494	0.403	0.494	0.433	0.647	0.647	0.545	0.540	0.552
s23	0.941	0.854	0.945	0.892	0.955	0.955	0.951	0.951	0.952
s25	0.934	0.952	0.937	0.951	0.949	0.949	0.939	0.940	0.948
s31	0.913	0.895	0.924	0.893	0.925	0.925	0.913	0.912	0.913
s32	0.953	0.930	0.954	0.936	0.960	0.960	0.953	0.953	0.955
s35	0.716	0.621	0.740	0.789	0.791	0.790	0.694	0.730	0.708
s38	0.935	0.949	0.934	0.948	0.953	0.953	0.935	0.935	0.939
s43	0.923	0.935	0.924	0.939	0.936	0.936	0.923	0.922	0.933
s49	0.896	0.913	0.896	0.907	0.900	0.900	0.896	0.895	0.897
s54	0.836	0.877	0.836	0.884	0.866	0.866	0.835	0.837	0.862
ALL	0.958	0.958	0.957	0.957	0.965	0.965	0.955	0.958	0.958

Subj. no	233	236	241	242	243	246	261	262 asym	262 amb
s2	0.849	0.845	0.848	0.849	0.848	0.862	0.851	0.850	0.852
s3	0.933	0.935	0.937	0.944	0.944	0.945	0.947	0.947	0.947
s8	0.900	0.904	0.843	0.843	0.845	0.847	0.930	0.933	0.929
s9	0.841	0.860	0.868	0.869	0.872	0.888	0.935	0.934	0.941
s11	0.941	0.949	0.944	0.947	0.946	0.952	0.965	0.965	0.968
s16	0.802	0.864	0.904	0.906	0.912	0.926	0.950	0.952	0.958
s20	0.942	0.953	0.918	0.919	0.933	0.945	0.959	0.962	0.963
s21	0.549	0.673	0.683	0.704	0.707	0.806	0.885	0.884	0.884
s23	0.952	0.961	0.870	0.879	0.901	0.932	0.969	0.968	0.969
s25	0.941	0.957	0.951	0.951	0.951	0.953	0.970	0.973	0.971
s31	0.924	0.926	0.897	0.896	0.895	0.899	0.920	0.920	0.923
s32	0.954	0.962	0.952	0.955	0.955	0.959	0.978	0.979	0.979
s35	0.745	0.797	0.620	0.622	0.786	0.788	0.774	0.781	0.787
s38	0.935	0.954	0.961	0.961	0.962	0.964	0.971	0.973	0.971
s43	0.923	0.935	0.944	0.946	0.948	0.952	0.967	0.966	0.970
s49	0.896	0.899	0.925	0.924	0.924	0.936	0.928	0.929	0.933
s54	0.836	0.865	0.898	0.897	0.891	0.894	0.914	0.924	0.917
ALL	0.960	0.965	0.961	0.897	0.891	0.894	0.974	0.976	0.976

Subj. no	263	266	611	621 SMvL	621 SvML	621 amb	631	661
s2	0.850	0.864	0.929	0.929	0.934	0.929	0.941	0.943
s3	0.951	0.955	0.972	0.975	0.975	0.973	0.975	0.977
s8	0.938	0.941	0.974	0.976	0.974	0.974	0.976	0.975
s9	0.941	0.950	0.979	0.979	0.984	0.978	0.987	0.988
s11	0.971	0.974	0.968	0.969	0.970	0.970	0.973	0.984
s16	0.953	0.959	0.968	0.968	0.977	0.967	0.978	0.979
s20	0.962	0.962	0.963	0.962	0.962	0.961	0.962	0.961
s21	0.885	0.900	0.968	0.970	0.968	0.967	0.971	0.976
s23	0.969	0.970	0.970	0.970	0.982	0.969	0.982	0.982
s25	0.974	0.974	0.966	0.966	0.972	0.967	0.972	0.974
s31	0.933	0.933	0.966	0.966	0.966	0.966	0.965	0.967
s32	0.981	0.982	0.983	0.984	0.984	0.985	0.985	0.985
s35	0.869	0.772	0.934	0.933	0.935	0.937	0.934	0.939
s38	0.975	0.975	0.978	0.979	0.978	0.978	0.981	0.984
s43	0.967	0.971	0.971	0.970	0.972	0.971	0.972	0.973
s49	0.929	0.937	0.968	0.968	0.968	0.969	0.969	0.977
s54	0.923	0.926	0.881	0.894	0.884	0.904	0.893	0.922
ALL	0.980	0.981	0.994	0.995	0.994	0.994	0.996	0.997

X. Appendix 4: Adjusted R^2 data for NP/Z fits

Subj. no	111	666	311	312	313 asym	313 length	314	316	321 closure
s2	0.557	0.959	0.861	0.859	0.870	0.866	0.869	0.869	0.860
s3	0.598	0.966	0.925	0.927	0.923	0.929	0.925	0.928	0.936
s8	0.503	0.981	0.943	0.945	0.943	0.944	0.945	0.944	0.943
s9	0.755	0.987	0.980	0.980	0.980	0.981	0.980	0.981	0.980
s10	0.833	0.973	0.872	0.880	0.870	0.873	0.880	0.879	0.938
s11	0.631	0.971	0.935	0.939	0.942	0.934	0.942	0.943	0.935
s15	0.623	0.960	0.839	0.837	0.837	0.871	0.835	0.881	0.840
s20	0.795	0.971	0.896	0.903	0.895	0.913	0.905	0.921	0.925
s21	0.481	0.962	0.913	0.914	0.913	0.914	0.916	0.916	0.917
s23	0.710	0.969	0.917	0.918	0.919	0.916	0.919	0.920	0.923
s25	0.671	0.988	0.969	0.977	0.969	0.970	0.976	0.979	0.982
s31	0.766	0.972	0.931	0.931	0.937	0.932	0.936	0.936	0.931
s32	0.718	0.985	0.954	0.954	0.953	0.953	0.953	0.952	0.962
s35	0.746	0.977	0.912	0.912	0.911	0.918	0.911	0.929	0.912
s38	0.733	0.978	0.977	0.977	0.978	0.977	0.977	0.977	0.977
s49	0.911	0.978	0.951	0.954	0.952	0.950	0.953	0.952	0.953
s51	0.868	0.986	0.962	0.963	0.972	0.973	0.980	0.980	0.964
s54	0.736	0.939	0.855	0.853	0.856	0.862	0.856	0.858	0.856
ALL	0.769	0.996	0.985	0.985	0.985	0.985	0.985	0.985	0.986

Subj. no	321 SMvL	321 SvML	322 closure	323 closure	324 closure	326 closure	331 asym	331 length	332 asym
s2	0.861	0.889	0.859	0.872	0.871	0.871	0.866	0.899	0.865
s3	0.924	0.940	0.938	0.935	0.937	0.939	0.924	0.942	0.925
s8	0.943	0.943	0.948	0.942	0.948	0.947	0.943	0.942	0.945
s9	0.980	0.982	0.980	0.980	0.979	0.981	0.981	0.981	0.980
s10	0.889	0.888	0.950	0.937	0.949	0.951	0.870	0.893	0.878
s11	0.934	0.934	0.939	0.941	0.943	0.943	0.945	0.934	0.946
s15	0.948	0.886	0.841	0.839	0.839	0.883	0.839	0.950	0.837
s20	0.938	0.938	0.937	0.931	0.936	0.949	0.897	0.949	0.908
s21	0.915	0.913	0.916	0.918	0.917	0.918	0.914	0.916	0.917
s23	0.918	0.917	0.923	0.924	0.925	0.926	0.926	0.917	0.925
s25	0.969	0.970	0.982	0.982	0.982	0.986	0.970	0.971	0.977
s31	0.931	0.941	0.931	0.937	0.937	0.937	0.947	0.946	0.947
s32	0.954	0.956	0.971	0.962	0.971	0.971	0.953	0.956	0.953
s35	0.951	0.922	0.911	0.911	0.910	0.929	0.912	0.951	0.912
s38	0.977	0.977	0.978	0.978	0.978	0.978	0.978	0.977	0.978
s49	0.950	0.953	0.953	0.953	0.952	0.952	0.957	0.954	0.958
s51	0.965	0.975	0.964	0.979	0.981	0.982	0.972	0.975	0.978
s54	0.855	0.862	0.864	0.856	0.868	0.871	0.859	0.863	0.860
ALL	0.987	0.989	0.987	0.872	0.988	0.988	0.986	0.989	0.986

Subj. no	333 asym	333 length	334 asym	336 asym	341	342	343 asym	343 length	344
s2	0.867	0.867	0.866	0.867	0.868	0.868	0.870	0.868	0.869
s3	0.923	0.929	0.924	0.927	0.935	0.937	0.933	0.936	0.936
s8	0.942	0.944	0.945	0.943	0.943	0.947	0.942	0.943	0.947
s9	0.981	0.981	0.981	0.982	0.981	0.980	0.981	0.981	0.980
s10	0.868	0.871	0.880	0.880	0.937	0.949	0.937	0.938	0.949
s11	0.945	0.948	0.946	0.947	0.945	0.949	0.944	0.948	0.948
s15	0.836	0.879	0.835	0.877	0.841	0.842	0.838	0.878	0.839
s20	0.896	0.920	0.906	0.922	0.934	0.940	0.931	0.944	0.940
s21	0.913	0.914	0.916	0.916	0.920	0.919	0.920	0.921	0.919
s23	0.927	0.930	0.927	0.928	0.929	0.931	0.930	0.932	0.931
s25	0.969	0.971	0.977	0.980	0.982	0.982	0.982	0.985	0.982
s31	0.946	0.946	0.946	0.946	0.948	0.949	0.947	0.947	0.949
s32	0.954	0.953	0.953	0.953	0.961	0.970	0.962	0.962	0.970
s35	0.910	0.922	0.910	0.929	0.911	0.911	0.910	0.925	0.909
s38	0.978	0.978	0.978	0.977	0.979	0.979	0.979	0.979	0.979
s49	0.958	0.957	0.959	0.958	0.957	0.957	0.958	0.957	0.958
s51	0.973	0.978	0.981	0.981	0.983	0.984	0.982	0.983	0.984
s54	0.860	0.862	0.859	0.861	0.858	0.867	0.859	0.863	0.866
ALL	0.985	0.986	0.985	0.986	0.988	0.989	0.988	0.988	0.989

Subj. no	346	361	362	363 asym	363 length	364	366	411	412
s2	0.868	0.899	0.902	0.902	0.908	0.904	0.921	0.886	0.885
s3	0.938	0.943	0.946	0.942	0.946	0.945	0.944	0.933	0.933
s8	0.946	0.943	0.947	0.941	0.943	0.946	0.945	0.960	0.969
s9	0.982	0.983	0.983	0.983	0.983	0.983	0.983	0.980	0.980
s10	0.950	0.941	0.954	0.941	0.945	0.954	0.953	0.966	0.967
s11	0.948	0.948	0.952	0.948	0.948	0.952	0.951	0.946	0.956
s15	0.878	0.959	0.959	0.955	0.963	0.959	0.962	0.845	0.845
s20	0.954	0.961	0.966	0.958	0.966	0.967	0.967	0.940	0.940
s21	0.919	0.928	0.928	0.928	0.927	0.926	0.930	0.929	0.929
s23	0.932	0.935	0.938	0.937	0.938	0.938	0.945	0.928	0.927
s25	0.985	0.984	0.984	0.984	0.985	0.984	0.985	0.983	0.983
s31	0.949	0.950	0.952	0.949	0.950	0.951	0.952	0.937	0.939
s32	0.970	0.962	0.971	0.963	0.963	0.971	0.971	0.973	0.975
s35	0.928	0.969	0.969	0.968	0.971	0.968	0.971	0.911	0.911
s38	0.979	0.979	0.979	0.978	0.978	0.979	0.979	0.978	0.979
s49	0.958	0.958	0.958	0.958	0.960	0.959	0.960	0.955	0.960
s51	0.984	0.985	0.982	0.986	0.984	0.984	0.984	0.964	0.964
s54	0.869	0.862	0.873	0.863	0.868	0.872	0.871	0.901	0.901
ALL	0.990	0.990	0.991	0.990	0.991	0.992	0.992	0.991	0.992

Subj. no	413 asym	414	421	431a	441	442	443 asym	444 asym	446
s2	0.895	0.894	0.887	0.886	0.891	0.890	0.893	0.891	0.891
s3	0.931	0.931	0.937	0.936	0.935	0.936	0.934	0.935	0.938
s8	0.960	0.970	0.968	0.969	0.970	0.970	0.970	0.970	0.970
s9	0.980	0.980	0.980	0.980	0.981	0.981	0.981	0.980	0.982
s10	0.966	0.967	0.969	0.970	0.970	0.969	0.969	0.969	0.969
s11	0.953	0.958	0.955	0.960	0.960	0.962	0.960	0.962	0.963
s15	0.842	0.842	0.845	0.844	0.844	0.843	0.841	0.840	0.878
s20	0.940	0.940	0.940	0.941	0.942	0.943	0.943	0.945	0.956
s21	0.929	0.929	0.929	0.930	0.929	0.931	0.929	0.930	0.931
s23	0.930	0.929	0.927	0.931	0.936	0.936	0.937	0.937	0.938
s25	0.983	0.983	0.983	0.984	0.983	0.984	0.983	0.984	0.988
s31	0.943	0.943	0.938	0.943	0.953	0.954	0.952	0.953	0.954
s32	0.973	0.975	0.973	0.973	0.972	0.977	0.973	0.977	0.977
s35	0.910	0.910	0.911	0.911	0.911	0.910	0.909	0.908	0.927
s38	0.979	0.979	0.978	0.980	0.979	0.980	0.979	0.979	0.979
s49	0.956	0.959	0.964	0.963	0.966	0.966	0.967	0.967	0.966
s51	0.974	0.980	0.964	0.982	0.983	0.980	0.984	0.983	0.985
s54	0.903	0.908	0.900	0.909	0.910	0.913	0.911	0.912	0.916
ALL	0.991	0.992	0.992	0.992	0.992	0.992	0.992	0.992	0.993

Subj. no	461	466	611	612 closure	613 asym	614	616	621 SMvL	621 SvML
s2	0.909	0.922	0.947	0.947	0.958	0.957	0.957	0.947	0.947
s3	0.945	0.942	0.958	0.958	0.958	0.958	0.958	0.960	0.964
s8	0.970	0.970	0.966	0.974	0.966	0.977	0.976	0.974	0.975
s9	0.983	0.983	0.980	0.980	0.980	0.982	0.982	0.981	0.983
s10	0.971	0.971	0.971	0.972	0.971	0.972	0.973	0.971	0.973
s11	0.963	0.965	0.953	0.964	0.961	0.966	0.966	0.955	0.953
s15	0.959	0.962	0.968	0.968	0.968	0.970	0.970	0.971	0.967
s20	0.968	0.972	0.964	0.965	0.965	0.965	0.964	0.965	0.965
s21	0.941	0.947	0.961	0.961	0.962	0.962	0.962	0.961	0.961
s23	0.940	0.947	0.946	0.946	0.949	0.953	0.953	0.945	0.947
s25	0.985	0.987	0.983	0.983	0.983	0.987	0.987	0.985	0.983
s31	0.954	0.954	0.954	0.956	0.961	0.960	0.960	0.956	0.955
s32	0.973	0.977	0.975	0.977	0.975	0.977	0.977	0.977	0.978
s35	0.969	0.971	0.974	0.975	0.975	0.976	0.976	0.975	0.974
s38	0.979	0.978	0.978	0.979	0.979	0.979	0.979	0.978	0.978
s49	0.966	0.967	0.962	0.968	0.964	0.967	0.967	0.962	0.962
s51	0.985	0.984	0.968	0.967	0.977	0.984	0.984	0.969	0.974
s54	0.916	0.919	0.923	0.924	0.926	0.935	0.935	0.924	0.934
ALL	0.994	0.995	0.994	0.995	0.994	0.995	0.995	0.994	0.994

Subj. no	631	641	661
s2	0.949	0.955	0.954
s3	0.963	0.961	0.965
s8	0.977	0.975	0.980
s9	0.983	0.981	0.986
s10	0.973	0.974	0.974
s11	0.954	0.968	0.969
s15	0.972	0.970	0.976
s20	0.965	0.967	0.968
s21	0.961	0.962	0.961
s23	0.947	0.956	0.968
s25	0.985	0.983	0.987
s31	0.960	0.971	0.971
s32	0.978	0.975	0.979
s35	0.976	0.976	0.978
s38	0.978	0.979	0.979
s49	0.962	0.974	0.976
s51	0.974	0.987	0.987
s54	0.937	0.934	0.937
ALL	0.995	0.996	0.996