

Running Head: FORGETTING IN WORKING MEMORY

Exploring the Forgetting Mechanisms in Working Memory:

Evidence from a Reasoning Span Test

Satoru Saito

Kyoto University, Japan

Christopher Jarrold

University of Bristol, UK

Deborah M Riby

Newcastle University, UK

Corresponding Author:

Satoru SAITO

Department of Cognitive Psychology in Education

Graduate School of Education

Kyoto University

Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501

JAPAN

Telephone: +81 75 753 3066

Facsimile: +81 75 753 3066

E-mail: S.Saito@mbox.kudpc.kyoto-u.ac.jp

Abstract

In working memory (WM) span tests participants have to maintain to-be-remembered information while processing other, potentially distracting, information. Previous studies have shown that WM span scores are greater when span lists start with a long processing task and end with a short processing than when these processing tasks are presented in the reverse order (e.g., Towse, Hitch, & Hutton, 2000). In Experiment 1, we obtained a similar stimulus order effect in a reasoning span test, using reasoning sentences that were equated for length in terms of the number of constituent words but which differed in processing complexity; span scores were greater when lists began with a complex sentence and ended with a simple sentence compared to when this stimulus order was reversed. In Experiment 2, the stimulus order effect was not found when processing duration was held constant while sentence complexity was varied using a computer-paced moving window presentation paradigm. These results suggest that duration-based constraints can affect degree of forgetting independently of the load generated during processing phases in WM span performance, and therefore imply that time-related forgetting can occur in WM span tests, particularly when the difficulty of the processing component blocks active maintenance of to-be-remembered material.

Key words: Working memory; reading span; task switching; resource sharing; short-term forgetting; stimulus order effect.

Exploring the Forgetting Mechanisms in Working Memory:

Evidence from a Reasoning Span Test

Working memory (WM) is a function that supports temporary storage of information in the service of goal-directed behaviour and higher cognitive processes (Baddeley, 2007). It is generally accepted that one of the most important characteristics of WM is that it is always required during performance of a target cognitive processing task, and could therefore be subject to distraction from these processing activities. Indeed, the development of WM span tests was essentially based on this approach to the measurement of measuring WM capacity (Case, Kurland, & Goldberg, 1982; Daneman & Carpenter, 1980), that is, in WM span tests participants have to maintain to-be-remembered information while processing other potentially distracting information. Although there are other measures which may tap WM capacity that do not contain an explicit processing component (Unsworth & Engle, 2007a), a large body of research has shown that processing plays a crucial role in disrupting maintenance in most WM tasks. However, what is less clear is the precise mechanism by this disruption occurs, or, in other words, what causes forgetting from WM. This issue is the focus of the current paper.

According to one account – the task switching hypothesis – the processing and maintenance aspects of WM span tests are performed sequentially and independently of each other. Towse and colleagues (Towse, Hitch, & Hutton, 1998; 2000) have argued that performance on WM span tests can therefore be explained by assuming that the to-be-remembered items are subject to temporal decay during the processing phase of the test list (see, Hitch, Towse, & Hutton, 2001 who adopted the mechanistically neutral term – *time-based forgetting*). As a consequence, the longer the participant is engaged in processing, the greater the extent of forgetting that occurs.

Potential evidence for time-based forgetting in WM comes from experiments that have manipulated retention requirements while attempting to hold constant overall processing demand. Towse et al. (1998) modified a standard reading span test, in which participants are required to read aloud a series of unrelated sentences (processing) and to recall the final word of each

sentence in the order in which they had occurred (storage), by using mixed lists of short and long sentences. In a Long-Final condition, a short sentence was presented at the first position in the list and a long sentence was presented at the final position. In a Short-Final condition, a long sentence was presented first and a short sentence was presented last in the list. A key strength of this manipulation is that it ensured that these two conditions differed only in the sentence order within a list and not in terms of to-be-remembered target words, sentences processed, total resources required for each list, and total duration of the test. Towse et al. (1998) assumed that if time-based forgetting of the to-be-remembered words is a determinant of reading span scores, then scores in the Short-Final condition should be higher than those in the Long-Final condition because the temporal duration of the final sentence determines the retention time of most of the to-be-remembered words within a list. Indeed, they found this sentence order effect (more generally, the stimulus order effect; see Hitch et al., 2001) among children (Towse et al., 1998) and in adults (Towse et al., 2000).

However, in a list with the Long-Final structure, most of the to-be-remembered items are exposed not only to a longer retention time but also to much more interference from processing of the final sentence than in a Short-Final list. A difference in interference between the two list structures could lead to differential span scores, irrespective of the difference in time-based forgetting within the lists (cf., Barrouillet, & Camos, 2001). Indeed, Saito and Miyake (2004) argued that representation-based interference, rather than time-based forgetting, might be the main mechanism of forgetting during reading span performance. They found that the stimulus order effect was observed even when the reading duration of the sentences was held constant and the amount of sentence processing required was varied (Experiment 3). In contrast, the stimulus order effect disappeared when reading duration was varied while holding constant the amount of sentence processing required (Experiments 4A & 4B). Thus, the amount of information processing of the final sentence is critical to the presence of the stimulus order effect in the reading span test. Consequently, they suggested that storage items are susceptible to interference from

representations generated by the processing phase (cf., the Feature Overwriting Hypothesis; Neath & Nairne, 1995; Oberauer & Lange, 2008), and that the degree of this interference will depend on the number of potentially interfering representations produced during processing rather than the time taken to complete processing operations (see also Maehara & Saito, 2007).

This representation-based interference account is well placed to explain the results of the existing studies on the stimulus order effect. However, it struggles to account for a finding from a study of children's WM performance carried out by Towse and Hitch (1995). They asked children to complete counting span tasks (cf. Case et al., 1982) in which they had to count the number of targets shown on successive cards while retaining the total of each counting operation for subsequent serial recall. In a 'feature' condition, the children counted blue target squares, which were shown among orange distractor triangles. In a 'conjunction' condition, the task was identical except that the target squares and non-target triangles were of the same colour (i.e., blue), which made the counting task more difficult and consequently made the processing duration longer. Towse and Hitch found that counting span scores were lower in the conjunction than in the feature condition, which was consistent with their suggestion that longer processing times induce a longer retention interval for memory items, leading to lower memory performance. This difference in span between the feature and the conjunction conditions is not consistent with predictions from the representation-based interference account because there is no reason to predict that the conjunction condition generates more representations than the feature condition.

One way of accounting for this discrepancy is to suggest that the degree of representation-based interference that occurs in a WM test is dependent on the participant's attempts to keep to-be-remembered information active in WM. For example, Jarrold and Bayliss (2007) argued that when a processing task embedded within a WM span test is not demanding enough to completely block the re-activation of memory items, then amount of forgetting will depend on the difficulty of the processing task, because this determines the time available for reactivation (cf. the time-based resource sharing or TBRS model; Barrouillet, Bernardin, & Camos, 2004). Conversely,

when a processing task is demanding enough to prevent re-activation processes, extent of forgetting will depend on the length of overall retention durations (cf. Lépine Bernadin, & Barrouillet, 2005). Unsworth and Engle (2007b) have proposed a similar theoretical framework, suggesting that WM span performance depends on the balance between two processes; attention-based active maintenance and cue-based retrieval from long-term memory (in their terms, primary memory and secondary memory, respectively). They suggest that in a WM task participants will attempt to keep memory items active in primary memory. However, when these items are displaced from primary memory, participants are forced to retrieve them from secondary memory, which they do on the basis of on temporal-contextual cues or temporal distinctiveness of the memory items, and thus sensitive to manipulations with temporal factors (cf. Glenberg & Swanson, 1986; Neath & Crowder, 1990; see also Nairne 2002). Indeed, potential measures of secondary memory, such as free recall tasks, show strong correlations with more standard WM span tests (Unsworth, 2007a). Other evidence for Unsworth and Engle's two-component framework comes from a variety of sources (see Unsworth & Engle, 2007b), including list length effects (Unsworth & Engle, 2006a) and differential patterns of errors (Unsworth & Engle, 2006b) in short-term memory and WM tasks.

This two component framework can explain both Saito and Miyake's (2004) evidence of representation-based interference and Towse and Hitch's (1995) finding of time-based effects, provided one assumes that the processing operations employed by Saito and Miyake allowed adults the opportunity to maintain items actively in WM while those employed by Towse and Hitch prevented ongoing active maintenance in children, thereby forcing a reliance on temporally-cued long-term memory. Although obviously post-hoc, this assumption is not unreasonable given other evidence that children find interleaved processing relatively more distracting than adults, perhaps because they are less efficient at engaging in active maintenance strategies (e.g., Cowan et al., 2005).

The aim of the current study was to provide a direct test of these suggestions, by asking

whether WM span performance is negatively affected by the temporal duration of processing phases when the processing component of these tasks is difficult enough to prevent active maintenance during the processing period. For this purpose, we developed a novel WM span test - the *reasoning span task*- in which a processing episode consisted of a verbal reasoning problem, and a storage episode was a single unrelated word. In the verbal reasoning problem a sentence and a pair of letters were presented (e.g., “A is followed by B – BA”) and participants were required to judge whether the order of letters shown to the right of the sentence was correctly represented by the sentence (see, Baddeley, 1968; Baddeley & Hitch, 1974; Hitch & Baddeley, 1976). This processing component was adopted because of existing evidence showing that the demands of performing these verbal reasoning problems are considerable even for adult participants. Hitch and Baddeley (1976) argued that in a preload situation, adult participants tended to refresh memory items before verifying such sentences, and then returned to the memory items after making the verification response, because active maintenance of memory items during sentence verification was not possible. They also showed that adults’ recall of preloaded memory items declined with increasing response times to these interleaved verbal reasoning sentences.

The use of these verbal reasoning problems allowed for the easy manipulation of sentence complexity without changing apparent sentence length, that is, it was possible to vary internal processing load (i.e., the number of mental operations) or temporal duration of the processing, irrespective of the number of words presented in each sentence. In the current study, we used mixed lists of simple and complex sentences in the reasoning span test, in line with the sentence order manipulation employed by previous studies. Specifically, we prepared two list conditions; Complex-Final and Simple-Final lists, equivalent to the Long-Final and Short-Final lists in Towse et al. (1998). In a Complex-Final condition, a simple sentence was presented first in a list and a complex sentence was presented at the final position in the list. In a Simple-Final condition, a complex sentence was presented first and a simple sentence was presented last in the list. This allowed us to examine the stimulus order effect in our novel reasoning span task, and thereby to

test whether forgetting in a WM span task is driven by representation-based interference or by temporal factors when rehearsal is assumed to be prevented during processing.

In Experiment 1, we attempted to replicate the stimulus order effect in the reasoning span test, using a self-paced reading paradigm in which participants were allowed to solve reasoning problems at their own pace but as quickly and accurately as possible. By examining participants' response times and accuracy to the reasoning components of the task, we were able to test the assumption that this reasoning component adequately blocked rehearsal. If participants attempt to maintain memory items actively and continuously during the verbal reasoning components of the span test, then we would expect longer reasoning times and less accurate reasoning responses for the reasoning episodes at the end of a span list. This is because participants would have to retain the maximum number of memory items at this point in the trial, and consequently the lengthening of processing time and the decrement of processing accuracy due to interruptions for maintenance activity would be greatest in this instance. In contrast, if participants do not engage in active maintenance during the reasoning phase of the reasoning span test, then we would not observe such processing position effects on reasoning time and accuracy.

Experiment 2 then critically tested whether time-based forgetting occurs in the reasoning span test by examining whether a stimulus order effect is still observed when a computer-paced moving window presentation paradigm is employed (cf. Saito & Miyake, 2004). If the stimulus order effect is due to representational interference then one would expect to observe it even when the duration of processing is kept constant across conditions; in contrast, if the effect in the reasoning span task is driven by time-related factors one would only observe it when processing duration is varied across conditions.

Experiment 1

Method

Participants

Thirty-two undergraduate students from the University of Bristol were paid or received

course credits for their participation in this experiment. All of them were native speakers of English, and their age ranged from 18 to 36, with a mean of 20.3 years.

Materials

Verbal reasoning sentences, adapted from the materials employed in Baddeley and Hitch's (1974) study were used in the present experiments. They were selected from a set of 4 (2 by 2) types of verbal reasoning problems; affirmative vs. negative and active vs. passive (see, Table 1). The verb in each sentence was either "follow" or "precede". The pair of letters that was included in each sentence was randomly selected from a set of 11 letters; *C, D, F, K, J, H, L, N, R, W, X*. Half of the reasoning sentences were correctly represented by a letter pair placed at the right side of the sentences (e.g., "K is followed by L - KL"; see, "True Pair" in Table 1), but the remainder were not (e.g., "J is preceded by R - JR"; see, "False Pair" in Table 1). Participants were required to judge whether the letter order was "True" or "False". Pilot testing showed that the reasoning judgment was faster and more accurate for the affirmative sentence problems than for the negative sentence problems and that any effect of the active-passive contrast in reasoning time was weaker than the affirmative-negative effect.¹ For this reason, we regarded the affirmative sentences as simple sentences and the negative sentence as complex sentences. Simple and complex sentences contained the same number of active and passive constructions. To equate the number of words included in simple and complex sentences "does" was added in affirmative-active sentences (e.g., A does follow B).

The experimental trials were created from a total of 84 reasoning problems (48 simple and 36 complex sentence problems). These were grouped into 24 problem sets, 6 sets at each of the four set sizes (2, 3, 4, and 5). A 2-problem set included one simple and one complex reasoning problem, a 3-problem set had two simple and one complex problem, a 4-problem set had two simple and two complex problems, and a 5-problem set had three simple and two complex problems. Half of the problem sets at each set size served as the Simple-Final lists, creating a total of 12 lists which begun with complex problem(s) and ended with simple problem(s). The

remaining half of the problem sets served as the Complex-Final lists. These 12 lists began with simple problem(s) and ended with complex problem(s). The assignment of the problem sets to the Simple-Final versus Complex-Final lists was counterbalanced across participants.

In the reasoning span test employed here, the to-be-remembered target word was presented after the true-false judgment to each reasoning problem. Eighty-four high-frequency one-syllable words were selected from the CELEX database (Max Plank Institute for Psycholinguistics, Nijmegen, 1995). The mean of the frequency values was 1950.9 within the corpus of 17.9 million words. They were randomly grouped into 24 sets, 6 sets at each of the four set sizes (2, 3, 4, and 5). The assignment of the word sets to the Simple-Final versus Complex-Final lists was counterbalanced across participants.

Procedure

We tested participants individually. In each trial, participants were shown a series of verbal reasoning problems one at a time, each of which was followed by a target word. All the problems and the target words were presented in the center of a computer display. The participants' task was to solve each verbal reasoning problem silently as quickly and accurately as possible and to click, using a computer mouse, the "True" or "False" button presented under the reasoning problems on the computer display. In addition, they were required to read aloud the target word and to try to remember it. The target word was presented immediately after a participant's response to the reasoning problem for one second and then it was replaced with the next reasoning problem. After all the problems and the target words in a set were presented, a question mark appeared in the center of the display, prompting participants to orally recall as many target words as possible from the set in their presented order. In scoring the data, we did not take account of the recall order, an approach consistent with previous studies (e.g., the scoring of reading span in Shah & Miyake, 1996).

The 24 lists (6 at each set size) were presented in a pre-specified random order, i.e., the presentation of the two types of lists (Simple-Final and Complex-Final lists) was not blocked and

lists of different sizes appeared in an apparently random order, not in ascending order. Participants were informed of how many problems would be presented on a given trial prior to the presentation of the first problem in that list. Before they began the 24 test trials, participants were given 4 practice trials, one at each set size, presented in ascending order (i.e., from 2 to 5).

Results

Recall Data

Individuals' memory performance was estimated for the Simple-Final condition and the Complex-Final condition independently. The scoring method was the *total score*, defined as the total number of target words correctly recalled across all sentence sets presented in a condition (Friedman & Miyake, 2000; Turner & Engle, 1989). The possible range for the total score measure was 0 to 42 for each stimulus order type (Simple-Final or Complex-Final). Although all the following statistical analyses were conducted with the total score because the continuous nature of this measure is statistically preferable, we also provide the traditional *span score*, defined as the highest set size for which the participant was able to correctly recall all of the words from at least two out of the three sets, adding half a point to the score if the participant's recall was accurate on one of the three sets at the next set size.

Table 2 shows the mean total scores and the mean span scores for the Simple-Final and Complex-Final conditions, and indicates that scores were greater in the Simple-Final condition than in the Complex-Final condition, $F(1, 31) = 4.66$, $MSE = 7.095$, $p = .039$, $partial \eta^2 = .13$.

Reasoning Performance Data

Reasoning accuracy. The correct response rates were higher for the simple problems ($M = 0.856$, $SD = 0.150$) than for the complex problems ($M = 0.819$, $SD = 0.139$), $F(1, 31) = 5.63$, $MSE = 0.003$, $p = .024$. Table 3 presents the mean correct response rates of the first and final reasoning problems for two types of lists. A two-way Analysis of Variance (ANOVA) revealed a significant interaction between problem position (first vs. last) and stimulus order (Simple-Final vs. Complex-Final), $F(1, 31) = 10.95$, $MSE = 0.008$, $p = .002$. Analyses of simple main effects

showed that participants tended to solve the first problems in the Complex-Final lists (i.e., simple problems) more accurately than the first problems in the Simple-Final lists (i.e., complex problems), $F(1, 31) = 4.14$, $MSE = 0.009$, $p = .050$, and that they solved the last problems in the Simple-Final lists (i.e., simple problems) more accurately than the last problems in the Complex-Final lists (i.e., complex problems), $F(1, 31) = 7.06$, $MSE = 0.008$, $p = .012$. There was no significant average difference in correct response rates between the Simple-Final and the Complex-Final lists and between the first problems and the last problems $F_s(1, 31) < 1$.

Reasoning time data. Reasoning times were shorter for the simple problems ($M = 3835$ ms, $SD = 951$) than for the complex problems ($M = 4576$ ms, $SD = 1042$), $F(1, 31) = 93.55$, $MSE = 187688.8$, $p < .001$. Table 4 presents the mean reasoning times of the first and final reasoning problems for the two types of lists. The calculation of the mean reasoning times used here was based on only those verbal reasoning problems that were judged correctly by each participant, given the concern that incorrect responses may have resulted from only partial processing of the reasoning problems. A two-way ANOVA revealed a significant interaction between problem position (first vs. last) and stimulus order (Simple-Final vs. Complex-Final), $F(1, 31) = 91.68$, $MSE = 211986.4$, $p < .001$. Analyses of simple main effects showed that participants took significantly longer to solve the first problems in the Simple-Final lists (i.e., complex problems) than the first problems in the Complex-Final lists (i.e., simple problems), $F(1, 31) = 54.49$, $MSE = 181590.6$, $p < .001$, with a reasoning time difference of approximately 790 ms. Similarly, participants took significantly longer to solve the last problems in the Complex-Final lists (i.e., complex problems) than the last problems in the Simple-Final lists (i.e., simple problems), $F(1, 31) = 41.04$, $MSE = 232543.5$, $p < .001$, with a reasoning time difference of approximately 770 ms. The ANOVA also showed that there was no significant overall reasoning time difference between the Simple-Final and the Complex-Final lists, $F(1, 31) < 1$, but showed that the reasoning times for the last problems were significantly longer than the first problems, $F(1, 31) = 56.65$, $MSE = 364961.5$, $p < .001$.

Discussion

Analyses of the reasoning data showed that reasoning accuracy was almost equivalent for the reasoning problems at the start and at the end of each list. This contrasts with what would be predicted if participants were engaging in active maintenance during the processing phases of the task, because in this case memory load would increase throughout the list, leading to poorer reasoning performance for final problems. However, there was evidence that processing position affected reasoning speed, with longer reasoning times for the problems at the end of span lists than for the first problems in the lists. It is worth noting that this effect on processing time has been observed in recent studies examining processing speed data in other WM span tests, involving English-speaking (Friedman & Miyake, 2004) and Japanese-speaking (Saito & Miyake, 2004; Maehara & Saito, 2007) adult participants. Its presence in the current data, coupled with the clear absence of any comparable effect on reasoning accuracy, suggests that our participants attempted to refresh memory items just before starting the verbal reasoning, leading to an overall delay of the response, particularly at final list positions, but that they did not engage in active maintenance during performance of the reasoning task. Consequently, the data are consistent with the suggestion that individuals were unable to engage in ongoing active maintenance of the to-be-remembered storage items during the processing phase of the task.

The presence of a stimulus order effect on recall in the current data is consistent with the possibility that (temporal-contextual) cue-based retrieval underpins reasoning span performance. It should be noted that even if participants attempt to refresh memory items prior to beginning processing, we would still expect temporal distinctiveness to determine the degree of forgetting observed provided that participants are unable to rehearse these items *during* the processing phase. Although it is likely that rehearsing memory items within a list would change the temporal distinctiveness of these items (see, Ward, 2001), it is important to emphasise that only the final reasoning problem in a list affects the retention duration of memory items in our stimulus order manipulation, and that this is true even when participants successfully rehearse all previous

memory items in that list before processing the final reasoning problem. However, this experiment cannot conclusively separate processing load and processing duration explanations of the stimulus order effect, as load and duration are necessarily confounded in the form of self-paced design employed here. Consequently, in the second experiment we directly examined whether temporal factors affect the occurrence of the stimulus order effect in the reasoning span test.

Experiment 2

The main purpose of Experiment 2 was to test the stimulus order effect using a computer-paced paradigm (Saito & Miyake, 2004). Although the original paradigm employed in the reading span procedure presented successive segments of an individual sentence *one after another*, here we presented the segments of each reasoning sentence *cumulatively* at a predetermined pace. In this way, we were able to provide participants with all the words included in a problem sentence up to the point at which they attempted to make a reasoning judgment, thereby minimizing the memory load involved in solving the problem (see, Figure 1).

This experiment was designed to tease apart the effects of two possible sources of the stimulus order effect on reasoning span: The *retention time* of memory items required during the last reasoning problems within a set and the *processing load* involved in the last reasoning problems. We manipulated the processing load required while holding constant the reasoning sentence presentation time for each problem. This manipulation was achieved in a Duration-Constant condition, in which both simple and complex reasoning sentences were divided into five segments plus one segment for the subsequent letter pair. Each sentence segment was presented at the same rate so that participants were forced to read all sentences within the same time window. If the temporal dynamics within a list are the critical factor for the stimulus order effect in this task, then one would not expect to find a significant difference between spans for Simple-Final and for Complex-Final lists in this condition. In contrast, if the amount of information processing required is a constraining factor, then one would expect a stimulus order effect.

Note that in this Duration-Constant condition the current participants have more time for

each simple reasoning problem and less time for each complex problem than the participants did in Experiment 1. Thus one may argue that there might be a small amount of ‘spare time’ available in the fixed time window for a simple reasoning problem but not for a complex one, and that this will lead to a relative advantage for Simple-Final lists at recall. In order to reduce this possibility, we asked participants to read aloud the reasoning sentence during its cumulative presentation without falling behind the pace of computer. This procedure aimed to force them to synchronize their mental pace with the presentation of each segment of the reasoning sentence, which should minimize the amount of any spare time available. Indeed, Hudjetz and Oberauer (2007) have shown that paced reading aloud of material prevents articulatory subvocal rehearsal from taking place during WM span tasks. However, their data indicates that paced reading aloud does not prevent all forms of active maintenance, and they argued that participants are able to engage in rapid ‘reactivation’ of memory items even under these conditions. We acknowledge this possibility in the current study, but note that this would lead to better recall performance for the Simple-Final lists than for the Complex-Final lists in the Duration-Constant condition because there may be space available for reactivating memory items at the end of Simple-Final lists but not at the end of Complex-Final lists. Therefore this concern becomes relevant only if we observe a stimulus order effect in the Duration-Constant condition, but not if we do not find it in that condition.

Having said this, even if a stimulus order effect is not observed in this study when duration is held constant, it is still possible that the disappearance of the effect might be due to the employment of the computer-paced presentation paradigm. To allow for a clear interpretation of the stimulus order effects in the Duration-Constant condition, we constructed a second condition, namely a Duration-Varied condition, in which both retention time and problem complexity were varied as in Experiment 1. We expected to observe a clear stimulus order effect in this condition, given that both the amount and length of processing differed across the two stimulus order list types.

Method

Participants

Sixty-four participants were recruited for this experiment. They were undergraduate students from the University of Bristol who were paid or received course credits for the participation in this experiment, or visiting students at the open days at the University. The age of participants ranged from 17 to 30, with a mean of 19.8 years in the Duration-Constant condition ($N = 32$) and ranged from 17 to 26, with a mean of 19.8 years in the Duration-Varied condition ($N = 32$). All of the 64 participants were native speakers of English and none of them had participated in Experiment 1.

Design

The experiment used a two (stimulus order; Simple-Final vs. Complex-Final) by two (reasoning presentation duration; Duration-Constant vs. Duration-Varied) factorial design. The former was manipulated within-subjects and the latter was a between-subject manipulation.

Materials and Procedure

The reasoning problems were produced in the similar way to those in Experiment 1. In this experiment, however, each problem sentence was divided into five segments plus one letter-pair segment (Duration-Constant condition), or divided into a fixed number of shorter segments depending on its sentence complexity (Duration-Varied condition). In the latter condition, the complex sentences were divided into six segments plus one letter-pair segment and the simple sentences into four segments plus one letter-pair segment. In both duration conditions, individual sentences appeared on a computer display at a fixed rate of 700 ms per segment and then a letter pair was presented at the right side of the sentence for 1800 ms. Thus, the total presentation duration was 5300 ms for all sentences in the Duration-Constant condition (see, Figure 2). In the Duration-Varied condition, the presentation duration was 6000 ms for a complex sentence and 4600 ms for a simple sentence (see, Figure 3). We determined the presentation rate for this experiment on the basis of the reasoning time data in Experiment 1 and pilot testing.

The target words were selected from the CELEX database as in Experiment 1, but were not the same words as those used previously. There were two different sets of 84 high-frequency one-syllable words, set A and set B (Means of the frequency were 1948.2 and 1937.2 counts from the CELEX corpus). Half of the participants in each reasoning duration condition remembered the words from set A, and the other half were given words from set B.

The experimental procedure was as follows. The experimenter first pressed a key to inform participants of the number of problems to be presented on the forthcoming trial. When participants were ready to proceed, the experimenter pressed the key again. Immediately after this second key press, the first problem from that trial began to be displayed. With the exception of the first segment of the problem, all the other parts of that problem were not presented initially. Instead, after the presentation of the first sentence segment for 700 ms, the computer displayed the next segment for 700 ms while leaving the first segment still present. This cycle continued until the end of that sentence. After the 700-ms presentation of the final segment of a sentence, a letter pair was presented for 1800 ms at the right side of the sentence, which was still present on the computer display. Then, the sentence and the letter pair disappeared after the 1800 ms presentation of the letter pair. Immediately after the disappearance of the sentence and the letter pair, a target word was presented at the center of the display for 1000 ms. The target word was then followed by the presentation of the first segment of the next sentence. After the target word of the final sentence for that set was presented, a question mark appeared on the computer display, and participants were required to orally report all the target words they could remember from that sentence set in their presented order.

Participants were required to read aloud each sentence presented to them, but not the letter pair in each reasoning problem. They were asked to keep up with the pace of computer presentation and to finish reading each segment before the next segment appeared (i.e., within the 700 ms time limit). When the letter pair was presented, they silently judged the correctness of the order of the letters and made their response using a computer mouse as in Experiment 1, but with

the constraint that this response was made within the 1800 ms interval before the presentation of the to-be-remembered word. Then, they read the target word aloud within the one-second interval in which it was presented. Other methodological details were identical to those of Experiment 1.

Results

Recall Data

Table 5 presents mean reasoning span scores and total scores for the two types of stimulus order lists in each duration condition. A two-way ANOVA revealed that although main effects of stimulus order and reasoning duration were not significant, $F(1, 62) = 1.12$, $MSE = 6.26$, $p = .293$; $F(1, 62) < 1$, respectively, the interaction between the two factors was significant, $F(1, 62) = 5.11$, $MSE = 6.26$, $p = .027$. Analysis of simple main effects showed that participants' recall performance was significantly better for the Simple-Final lists than for the Complex-Final lists only in the Duration-Varied condition, $F(1, 31) = 5.10$, $MSE = 6.77$, $p = .031$, but not in the Duration-Constant condition, $F(1, 31) < 1$.

Although these data imply that temporal duration is the crucial factor underpinning the occurrence of the stimulus order effect in the reasoning span test, it is still possible to argue that solving complex problems does not necessarily generate more representations than solving simple problems in this span test. Indeed, it may be the case that participants did not attempt to solve the reasoning problems when they responded incorrectly, due to the difficulty of the reasoning required. To test this possibility, we analyzed recall performance by only using the lists in which participants solved the list final problems correctly. Because the number of such lists varied across conditions, we used proportion of correct recall scores for this analysis. The mean correct recall was .818 ($SD = .145$) and .812 ($SD = .110$) for Complex-Final and Simple-Final conditions, respectively, in the Duration-Constant condition, and .790 ($SD = .142$) and .852 ($SD = .101$) for Complex-Final and Simple-Final conditions, respectively, in the Duration-Varied condition. A two-way ANOVA revealed a significant interaction between stimulus order and reasoning duration factors, $F(1, 62) = 4.45$, $MSE = 0.008$, $p = .039$. Once again, the stimulus order effect

was observed only in the Duration-Varied condition but not in the Duration-Constant condition.

Reasoning Performance Data

Because presentation durations of the reasoning sentences and the response windows were fixed in this experiment and because the presentation of sentence segments was at a constant regular rate, we would expect participants to also make their responses at a regular pace. For this reason, we did not analyse reasoning times in this experiment. Rather, we report below analyses of the reasoning accuracy data only.

The correct response rates were higher for the simple problems ($M = .818$, $SD = .133$) than for the complex problems ($M = .677$, $SD = .153$), $F(1, 62) = 81.89$, $MSE = 0.008$, $p < .001$. Tables 6a and 6b present the mean correct response rates for the first and final reasoning problems of the two types of lists in Duration-Constant (Table 6a) and Duration-Varied conditions (Table 6b). A three-way ANOVA with duration condition (constant vs. varied), problem position (first vs. final), and stimulus order factors (Simple-Final vs. Complex-Final) revealed a significant interaction between problem position and stimulus order, $F(1, 62) = 66.10$, $MSE = 0.022$, $p < .001$. Analyses of simple main effects showed that participants solved the first problems in the Complex-Final lists (i.e., simple problems) more accurately than the first problems in the Simple-Final lists (i.e., complex problems), $F(1, 62) = 50.35$, $MSE = 0.020$, $p < .001$, and that they solved the last problems in the Simple-Final lists (i.e., simple problems) more accurately than the last problems in the Complex-Final lists (i.e., complex problems), $F(1, 62) = 26.29$, $MSE = 0.020$, $p < .001$. The main effect of problem position was also significant, $F(1, 62) = 9.50$, $MSE = 0.015$, $p = .003$, reasoning problems in the first position were solved more accurately than those in the final position. All other main effects and interactions were not significant.

One final point to note about the reasoning accuracy data is a suggestion that simple problems in the Duration-Constant condition were solved *more accurately* than those in the Duration-Varied condition (means collapsing serial position = .833 and .802, respectively) while complex problems in the Duration-Constant condition were solved *less accurately* than those in

the Duration-Varied condition (means collapsing serial position = .663 and .691, respectively).

This raises the possibility that the stimulus order effect on recall that is observed in the Duration-Varied condition reflects a trade-off between recall performance and reasoning accuracy that does not occur in the Duration-Constant condition. However, the differences of reasoning accuracy between the two duration conditions were not significant for the simple problems, $t(62) = 0.934$, $p = .354$, $d = .234$, and for the complex problems, $t(62) = 0.725$, $p = .471$, $d = .181$. Moreover, in order to secure enough power (i.e., .80) for detecting these small differences one would need 454 and 750 participants for the Duration-Constant and the Duration-Varied condition, respectively. In other words, there is no evidence to support the view that participants are prioritising reasoning accuracy differently in the two duration conditions.

Discussion

The stimulus order effect was observed in the Duration-Varied condition. However, this was not the case in the Duration-Constant condition of our reasoning span test. The absence of the stimulus order effect in this condition indicates that the factor driving this effect in the reasoning span test is not processing load (i.e., the number of mental operations or representations generated during processing), because this does not vary across the two duration conditions. Rather time-related forgetting appears to be the determinant of the stimulus order effect in this task.

Initial list problems were solved reliably more often than were final list problems. This "problem position" effect on reasoning accuracy was not observed in Experiment 1, although a similar position effect was observed in the reasoning time data in that experiment. The fact that a position effect was not observed on reasoning accuracy in Experiment 1 suggested that the observed effect on reasoning time in that experiment was due to participants delaying the start of processing activities in order to reactivate memory items, rather than any reflection of ongoing rehearsal during processing. If participants applied the same strategy in the current experiment, and delayed the onset of processing to any degree, then this would necessarily have an effect on processing accuracy because the time window for processing was limited in Experiment 2.

General Discussion

This study explored the nature of forgetting mechanisms underpinning WM performance by examining whether the stimulus order effect reported in previous studies is observed in our novel reasoning span test, and whether the effect is sensitive to an experimental manipulation of the temporal dynamics of the span lists. Experiment 1 showed a clear stimulus order effect; that is, Simple-Final lists led to better performance than Complex-Final lists. In Experiment 2, the stimulus order effect was observed only in the Duration-Varied condition in which both the processing load and the retention duration were varied, but not in the Duration-Constant condition in which we only varied the processing load and held the retention duration constant. Furthermore, an effect of problem position was found on reasoning time in Experiment 1, that is, the average reasoning time for the last problems in lists was significantly longer than that for the first problems, but this effect was not observed on reasoning accuracy in Experiment 1. In contrast, average correct reasoning rate for the last problems was significantly lower than that for the first problems in Experiment 2.

In the following sections, we first discuss the source of the processing position effect outlined above, before turning to an explanation of the stimulus order effects observed in our data, and finally concluding with general implications for the nature of forgetting observed across different WM span tasks.

Sources of the Processing Position Effect

Because the final problems, but not the first problems in a list, have to be solved in a situation in which participants might attempt to maintain the previously presented memory items, the presence of the processing position effect on reasoning time in Experiment 1 is consistent with the view that memory load affects processing performance within a WM span list (Saito & Miyake, 2004). However, this does not necessarily mean that participants were engaging in active maintenance *during* the processing activity. A slowing down of measured ‘processing time’ could equally reflect participants attempting to reactivate or encode memory items immediately before

starting the processing, as Hitch and Baddeley (1976) argued (this possibility has also been noted in previous studies of WM performance, Engle, Cantor, & Carullo, 1992; Friedman & Miyake, 2004; Hitch et al., 2001; Towse, Hitch, & Hutton, 2002).

Indeed, the fact that processing position had no effect on reasoning accuracy in Experiment 1 strongly suggests that participants were not engaging in reactivation or rehearsal while carrying out the reasoning problems employed in our task. An effect of processing position on reasoning accuracy was observed in Experiment 2, but we argue that this reflects the fact that a time limit was set for responses in this experiment. In other words, if participants attempt to reactivate or rehearse memory items prior to engaging in processing, this would necessarily limit their time for reasoning, particularly at later list positions, leading to an indirect effect on reasoning accuracy. Certainly, across the two experiments the pattern of response time and accuracy data from the processing component of our reasoning span task suggests that participants were not able to engage in active maintenance once they had begun processing the reasoning problems.

Sources of the Stimulus Order Effect

The fact that a stimulus order effect was observed only in the Duration-Varied condition and not in the Duration-Constant condition in Experiment 2 indicates that forgetting seems to occur as a function of retention duration within the WM span lists employed here. This result is consistent with the findings from Towse and colleagues' studies (Hitch et al., 2001; Towse et al., 1998, 2000) that suggest that temporal factors are an important determinant of WM performance. However, it appears not to be in line with the predictions of Barrouillet and colleagues' TBRS model (Barrouillet et al., 2004), which assumes that processing pace (defined as the number of processing actions per given time) determines the size of the processing load effects on recall. According to this model, participants use any available opportunities within the processing phase of a WM span task to engage in maintenance activities, and these opportunities are greater when processing is presented at a slower pace. Consequently, the TBRS model might be thought to

predict better memory performance for the Simple-Final lists than for the Complex-Final lists in the Duration-Constant condition because participants might be assumed to have more spare time available for maintenance activities at the end of Simple-Final lists.

However, the processing position effects on reasoning accuracy just discussed indicate that we were successful in our aim of preventing participants from engaging in any maintenance activities during the processing phase of our task. We would suggest that, by comparison, the processing tasks used in Barrouillet and colleagues' studies (Barrouillet et al., 2004; Barrouillet, Bernadin, Potrat, Vergauwe, & Camos, 2007) were relatively easy (producing a relatively slow processing pace), allowing participants to reactivate memory items *during* the processing phases of their tasks, leading to an effect of pace on performance (see Hudjetz & Oberauer, 2007). In contrast, if participants did not engage in active maintenance of memory items during performance of the reasoning tasks in the present study we would expect that memory items must be retrieved from long-term memory at recall. In this case, retrieval search processes would be based on temporal-contextual cues (e.g., Unsworth & Engle, 2006b), and recall levels should be affected by the retention duration of memory items and not by the pace of presentation of processing.

The suggestion that temporal factors affect adults' WM span performance when processing blocks maintenance activities is consistent with the results of Towse and Hitch's (1995) study, which showed that temporal factors constrain counting span performance in children who may not engage in active maintenance during processing (i.e., during counting). It can also explain the apparent contrast between the current findings and those of Saito and Miyake (2004), which led them to propose representation-based interference as a forgetting mechanism within WM span tests. In their Experiment 3, Saito and Miyake (2004) found a significant stimulus order effect in a reading span test when the reading duration of the sentences was held constant by the computer-paced moving window manipulation and the amount of sentence processing required was varied; the same paradigm was used in the current study, but clearly the opposite effects on

performance were observed (a stimulus order effect was seen only in the Duration-Varied condition).

As already noted, recent work by Hudjetz and Oberauer (2007) suggests that the processing demands employed by Saito and Miyake (2004) would have been sufficient to block subvocal articulatory rehearsal, but that participants would have still been able to engage in a more rapid reactivation of memory items while reading aloud the processing items. We assume that this ongoing reactivation was subject to distraction from representations generated during reading, causing representation-based interference. However, we again argue that the use in the current study of a more demanding processing task than simple reading aloud will have limited participants' ability to engage in any form of active maintenance while processing, thereby removing any representational-interference effects that might be caused by the overlap of storage and processing representations in WM.

The Nature of Forgetting in Working Memory Span Tests

Unsworth and Engle (2007) have argued that there are two factors that constrain WM span performance: active maintenance in primary memory and cue-based retrieval from secondary memory, with the latter being subject to general long-term memory phenomena (e.g., temporal distinctiveness; Glenberg & Swanson, 1986). The current and other related data support the presence and the distinction of these two influences on WM span performance. Our new proposal here is that these two factors are subject to two different *forgetting mechanisms*: Representation-based interference and time-related forgetting, respectively. If the processing activity embedded in any WM span task is not sufficiently demanding to prevent ongoing active maintenance of storage items during the processing phase of the task, then participants will attempt to maintain to-be-remembered information in this way. These active memory representations will therefore be subject to interference, and we would expect disruption when the processing activity generates representations that are similar to, and interfere with, them, with the degree of disruption depending on the number of interfering representations generated (Maehara & Saito, 2007; Saito

& Miyake, 2004). In addition, if active maintenance is possible within the processing phase of a WM task then one would expect to observe other manifestations of such activities, including the pace effects observed by Barrouillet and colleagues in their work (Barrouillet et al., 2004, 2007; see also Hudjetz & Oberauer, 2007).

In contrast, if active maintenance is not possible, then participants will be forced to retrieve to-be-recalled items from long-term memory on the basis of a search process that requires temporal-contextual cues. One type of cue-based retrieval model that directly refers to temporal aspects of cues is temporal distinctiveness theory (Glenberg & Swanson, 1986). According to this theory, recall performance is determined by item temporal distinctiveness, which is affected by two factors: the time elapsed since the target memory event and the time between competing events (see, Greene, 1992 for various theoretical discussions on this topic). Consequently recent events are more temporary distinctive than older events and one would therefore expect more accurate retrieval for memory items with shorter rather than longer retention durations.²

Of course, these two processes (active maintenance and cue-based retrieval) are not exclusive. If a participant fails to maintain memory items actively, these items might still be retrieved at recall. Therefore, recall performance in a WM span test with an easy processing task, which allows a degree of active maintenance of memory items, might be a consequence of successful active maintenance and successful retrieval of items that were not actively maintained. If this is the case, both representation-based interference and time-based forgetting could operate in such a WM span test. In contrast, when the processing task in a WM span test is difficult enough to completely prevent participants from active maintenance, memory items will be exclusively retrieved from long-term memory at recall. In this case, the forgetting mechanism that constrains memory performance appears to be time-related. Consequently, the key theoretical and practical implication of these arguments is that the nature of forgetting in WM varies depending on the difficulty of the processing task included in a WM span test, and the extent to which this prevents individuals from maintaining active representations of the to-be-remembered material.

References

- Baddeley, A. D. (1968). A 3 min reasoning test based on grammatical transformation. *Psychonomic Science*, 10, 341-342.
- Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *Psychology of learning and motivation VIII* (pp. 47-90). New York: Academic Press.
- Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, 133, 83-100.
- Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 570-585.
- Barrouillet, P., & Camos, V. (2001). Developmental increase in working memory span: Resource sharing or temporal decay? *Journal of Memory and Language*, 44, 1-20.
- Case, R., Kurland, M. D., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, 33, 386-404.
- Cowan, N., Elliot, E. M., Saults, J. S., Morey, C., Mattox, S., Hismjatullina, A., et al. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive abilities. *Cognitive Psychology*, 51, 42-100.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 972-992.
- Friedman, N. P., & Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. *Journal of Experimental Psychology: General*, 129,

61–83.

- Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, 51, 136–158.
- Glenberg, A. M., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 3–15.
- Green, R. L. (1992). *Human memory: Paradigms and paradoxes*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Hitch, G. J., & Baddeley, A. D. (1976). Verbal reasoning and working memory. *Quarterly Journal of Experimental Psychology*, 28, 603–621.
- Hitch, G. J., Towse, J. N., & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General*, 130, 184–198.
- Hudjetz, A., & Oberauer, K. (2007). The effects of processing time and processing rate on forgetting in working memory: Testing four models of the complex span paradigm. *Memory & Cognition*, 35, 1675–1684.
- Jarrold, C., & Bayliss, D. M. (2007). Variation in working memory due to typical and atypical development. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory*, pp. 134–161. New York: Oxford University Press.
- Lépine, R., Bernadin, S., & Barrouillet, P. (2005). Attention switching and working memory spans. *European Journal of Cognitive Psychology*, 17, 329–345.
- Lewandowsky, S., Brown, G. D. A., Wright, T., & Nimmo, L. M. (2006). Timeless memory: Evidence against temporal distinctiveness models of short-term memory for serial order. *Journal of Memory and Language*, 54, 20–38.
- Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in

- short-term serial recall. *Psychonomic Bulletin & Review*, 11, 771-790.
- Maehara, Y., & Saito, S. (2007). The relationship between processing and storage in working memory span: Not two sides of the same coin. *Journal of Memory and Language*, 56, 212-228.
- Max Plank Institute for Psycholinguistics, Nijmegen (1995). *The CELEX Lexical Database, release 2*.
- Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, 53, 53-81.
- Neath, I., & Crowder, R. G. (1990). Schedules of presentation and temporal distinctiveness in human memory. . . *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 316-327.
- Neath, I., & Nairne, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review*, 2, 429-441.
- Oberauer, K., & Lange, E. B. (2008). Interference in verbal working memory: Distinguishing similarity-based confusion, feature overwriting, and feature migration. *Journal of Memory and Language*, 58, 730-745.
- Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing-storage relationship in reading span performance. *Journal of Memory and Language*, 50, 425-443.
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, 125, 4-27.
- Towse, J. N., & Hitch, G. J. (1995). Is there a relationship between task demand and storage space in tests of working memory capacity? *Quarterly Journal of Experimental Psychology*, 48A, 108-124.
- Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A reevaluation of working memory capacity in

- children. *Journal of Memory and Language*, 39, 195–217.
- Towse, J. N., Hitch, G. J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition*, 28, 341–348.
- Towse, J. N., Hitch, G. J., & Hutton, U. (2002). On the nature of the relationship between processing activity and item retention in children. *Journal of Experimental Child Psychology*, 82, 156–184.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127–154.
- Unsworth, N. (2007). Individual differences in working memory capacity and episodic retrieval: Examining the dynamics of delayed and continuous distractor free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 1020–1034.
- Unsworth, N., & Engle, R. E. (2006a). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, 54, 68–80.
- Unsworth, N., & Engle, R. E. (2006b). A temporal–contextual retrieval account of complex span: An analysis of errors. *Journal of Memory and Language*, 54, 346–362.
- Unsworth, N., & Engle, R. E. (2007a). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, 133, 1038–1066.
- Unsworth, N., & Engle, R. E. (2007b). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114, 104–132.
- Ward, G. (2001). A critique of the working memory model. In J. Andrade (Ed.), *Working memory in perspective* (pp. 219–239). Hove: Psychology Press.

Author Notes

Satoru Saito, Department of Cognitive Psychology in Education, Kyoto University;
Christopher Jarrold, Department of Experimental Psychology, University of Bristol; Deborah M
Riby, Department of Psychology, Newcastle University. SS is a Visiting Fellow at the
Department of Experimental Psychology at the University of Bristol.

This research was supported by a Grant-in-Aid for Scientific Research (Project Number
14710082) from the Ministry of Education, Culture, Sports, Science, and Technology in Japan,
by the 21st Century COE Program (D-10 to Kyoto University), MEXT, Japan, by Canon
Foundation in Europe to Satoru Saito, and by a co-operative group component grant from the
United Kingdom Medical Research Council to Chris Jarrold and Alan Baddeley (grant number
G0000258, within co-operative group grant number G9901359). We thank Nelson Cowan for
helpful comments on an earlier version of this paper.

Correspondence concerning this article should be addressed to Satoru Saito, Department
of Cognitive Psychology in Education, Graduate School of Education, Kyoto University,
Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan. Electronic mail may be sent to
S.Saito@mbox.kudpc.kyoto-u.ac.jp

Footnotes

1. In Experiment 1, we obtained significant effects of the affirmative-negative contrast and of the active-passive contrast, perfectly replicating the results of Baddeley and Hitch (1974). The time difference of the effect was larger between the affirmative and the negative sentences (1867 ms) than between the active and the passive sentences (741 ms), supporting the validity of the selection of simple and complex problems in the current study.

2. Although there are concerns as to whether temporal distinctiveness or other temporal factors can explain data from immediate serial recall (see, Lewandowsky, Duncan, & Brown, 2004), these concerns do not preclude the possibility that temporal distinctiveness accounts for phenomena in longer-term situations like the continuous distractor paradigm (see, Lewandowsky, Brown, Wright, & Nimmo, 2006). It might be worth noting that the duration manipulation in our study cannot separate the temporal distinctiveness explanation from a temporal decay account because we manipulated only retention durations and not inter-item intervals. However we incline towards the temporal distinctiveness account to explain the stimulus order effect reported here for two reasons. First the phenomenon seems to occur in a long-term situation, where other evidence suggests that interference rather than temporal decay explanations are better placed to account for forgetting, and second because a temporal distinctiveness account might provide a better explanation of other related phenomenon (e.g., Unsworth & Engle, 2006a, 2006b).

Table 1

Examples of problem sentences used in the verbal reasoning task.

Complexity Category	Gramatical Form	Example of Sentence	True Pair	False Pair
Simple Problem	Affirmative-Active	D does follow X	XD	DX
	Affirmative-Passive	J is preceded by H	HJ	JH
Complex Problem	Negative-Active	L doesn't precede C	CL	LC
	Negative-Passive	N isn't followed by W	WN	NW

Table 2

Mean span and total scores for the two list types in Experiment 1.

Measure	Simple-Final List		Complex-Final List		Partial η^2
	M	SE	M	SE	
Span Scores (max. 5)	4.00	0.14	3.72	0.18	.09
Total Scores (max. 42)	36.31	0.68	34.88	0.89	.13

Table 3

Mean correct response rates (with standard errors in parentheses) at the first and final problem positions in each stimulus order condition in Experiment 1.

	First Problem	Final Problem	Mean
Simple-Final List	.810 (.033)	.859 (.030)	.835 (.028)
Complex-Final List	.857 (.031)	.799 (.031)	.828 (.028)
Mean	.833 (.030)	.833 (.028)	

Note. The calculation of the mean correct response rate was based on all the problems used in the reasoning span test.

Table 4

Mean reasoning times in ms (with standard errors in parentheses) at the first and final problem positions in each stimulus order condition in Experiment 1.

	First Problem	Final Problem	Mean
Simple-Final List	4070 (168)	4094 (192)	4082 (173)
Complex-Final List	3284 (134)	4867 (218)	4075 (162)
Mean	3677 (142)	4481 (196)	

Note. The calculation of the mean reasoning times was based on only those verbal reasoning problems that were correctly judged by each participant.

Table 5

Mean span and total scores for the two list types in each of presentation methods in

Experiment 2.

Presentation Condition	Measure	Simple-Final List		Complex-Final List		Partial η^2
		M	SE	M	SE	
Duration Constant	Span Scores (max. 5)	3.45	0.18	3.42	0.19	.00
	Total Scores (max. 42)	32.44	0.40	32.97	0.41	.02
Duration Varied	Span Scores (max. 5)	3.66	0.16	3.11	0.19	.22
	Total Scores (max. 42)	33.44	0.36	31.97	0.43	.14

Table 6a

Mean correct response rates (with standard errors in parentheses) for the two list types at the first and final problem positions in the Duration-Constant condition of Experiment 2.

	First Problem	Final Problem	Mean
Simple-Final List	.667 (.079)	.799 (.076)	.733 (.070)
Complex-Final List	.875 (.062)	.648 (.072)	.762 (.061)
Mean	.771 (.064)	.724 (.066)	

Table 6b

Mean correct response rates (with standard errors in parentheses) for the two list types at the first and final problem positions in the Duration-Varied condition of Experiment 2.

	First Problem	Final Problem	Mean
Simple-Final List	.708 (.076)	.786 (.073)	.747 (.070)
Complex-Final List	.852 (.067)	.680 (.080)	.766 (.067)
Mean	.780 (.065)	.733 (.071)	

Figure Captions

Fig. 1. Schematic illustration of the computer-paced presentation method used in Experiment 2. Each problem sentence is presented cumulatively at a fixed rate. Participants were requested to read sentence words aloud when they appear (except for the presentation of ":"). When a letter pair was presented at the right side of the sentence, which was still present on the computer display, participants make a reasoning judgment. The sentence and the letter pair disappeared after the 1800 ms presentation of the letter pair. A target word was then presented at the center of the display for 1000 ms at which point participants read the word aloud and attempted to remembered it.

Fig. 2. Schematic illustration of the experimental manipulation for the Duration-Constant condition in Experiment 2. In this condition, all the reasoning problems were gradually presented with a total presentation duration of 5300 ms regardless of problem complexity. Therefore the retention durations for the Simple-Final list and that for the Complex-Final list are the same.

Fig. 3. Schematic illustration of the experimental manipulation for the Duration-Varied condition in Experiment 2. In this condition, the complex reasoning problems were presented for 6000 ms in total and the simple reasoning problems were presented for 4600 ms in total. The retention duration is consequently longer for the Complex-Final list than for the Simple-Final list.

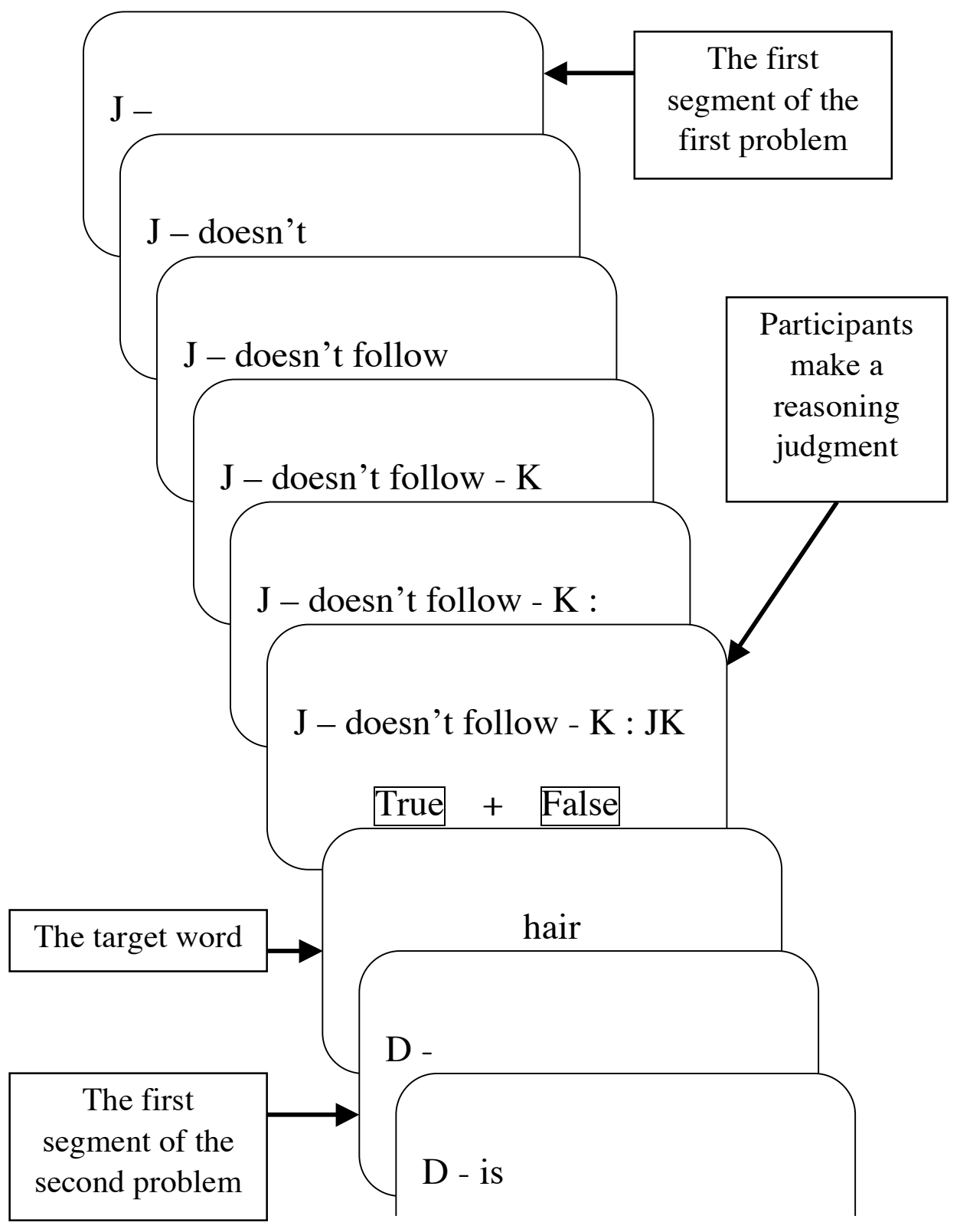


Fig. 1

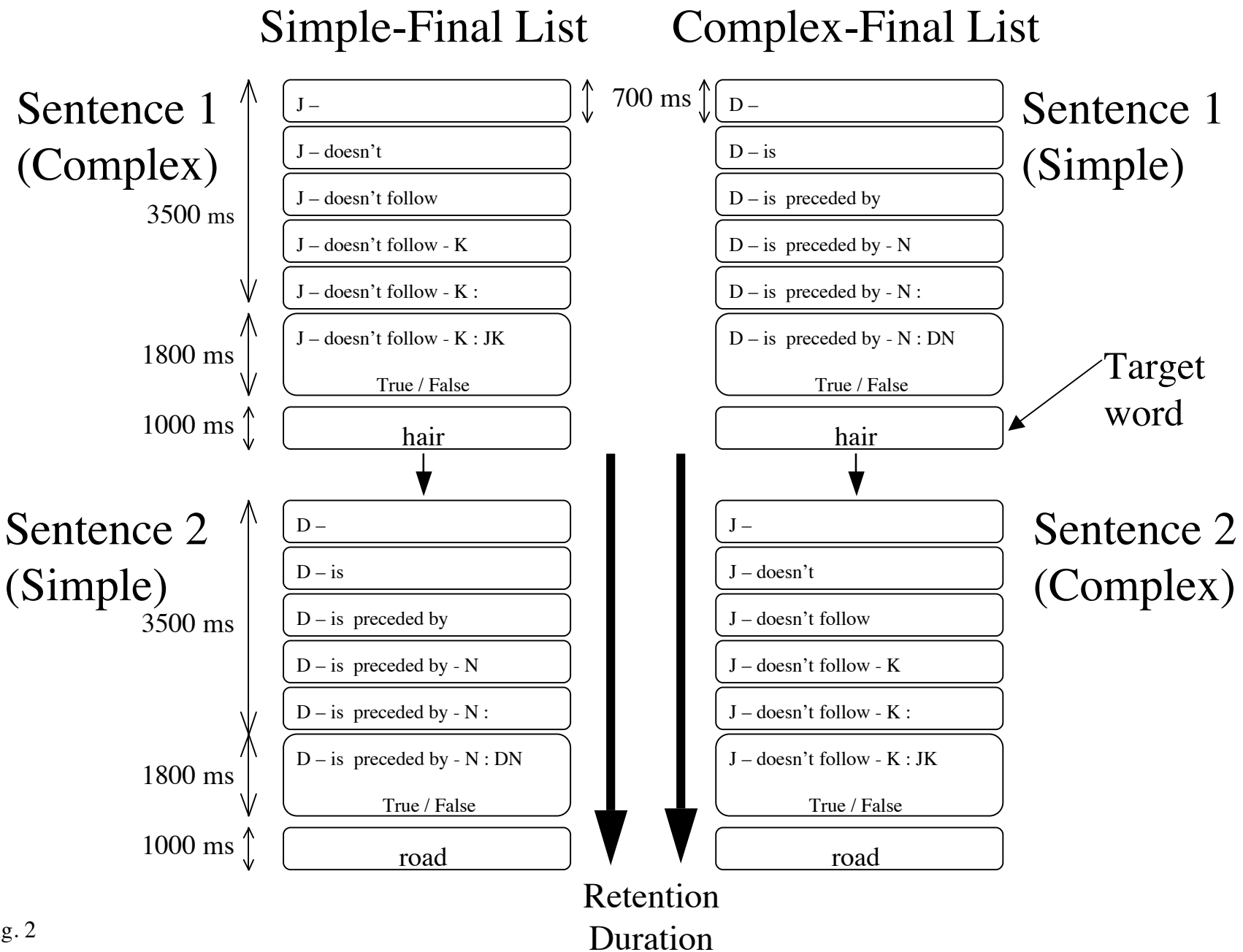


Fig. 2

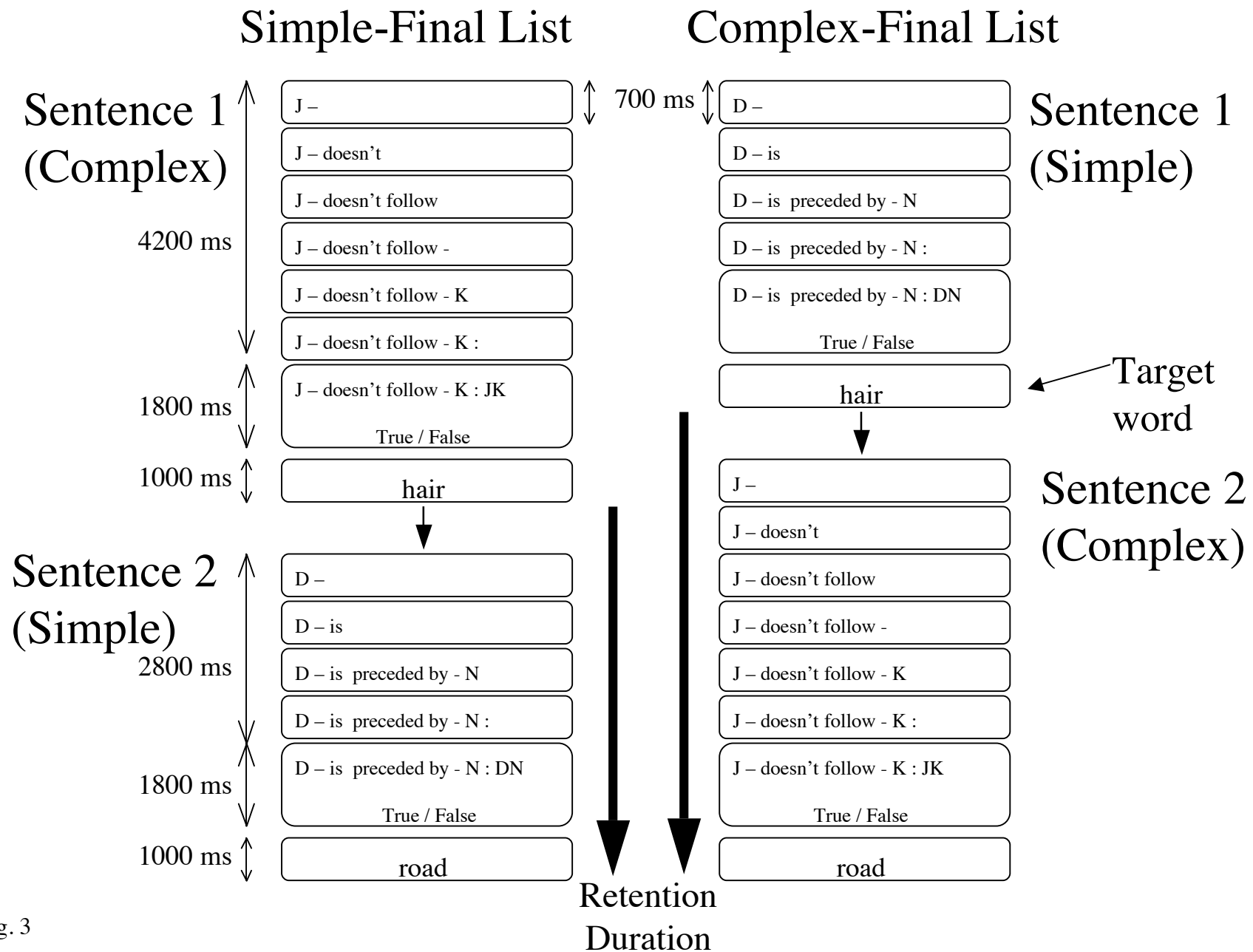


Fig. 3