Title

Investigating the inner speech of people who stutter: Evidence for (and against) the Covert Repair Hypothesis.

Authors

Paul H. Brocklehurst & Martin Corley
Department of Psychology.
School of Philosophy Psychology and Language Sciences.
The University of Edinburgh.
7 George Square,
Edinburgh.
EH8 9JZ
Scotland. UK.
P.H.Brocklehurst@sms.ed.ac.uk
Martin.Corley@ed.ac.uk

Abstract

In their Covert Repair Hypothesis, Postma and Kolk (1993) suggest that people who stutter make greater numbers of phonological encoding errors, which are detected during the monitoring of inner speech and repaired, with stuttering-like disfluencies as a consequence. Here, we report an experiment that documents the frequency with which such errors are made. Thirty-two people who stutter (PWS) and thirty-two normally-fluent controls, matched for age, gender and education, recited tonguetwisters and self-reported any errors they perceived themselves to have made. In 50% of trials the tonguetwisters were recited silently and errors reported were those detected in inner speech. Compared to controls, PWS produced significantly more word-onset and word-order errors. Crucially, this difference was found in inner as well as in overt speech. Comparison of experimenter ratings and participants' own self-ratings of their overt speech revealed similar levels of accuracy across the two groups, ruling out a suggestion that PWS were simply more sensitive to the errors they made. However, the frequency of participants' inner-speech errors was not correlated to their SSI4 scores, nor to two other measures of stuttering severity. Our findings support Postma and Kolk's contention that, when speech rate is held constant, PWS make, and therefore detect, more errors of phonological encoding. They do not, however, support the hypothesis that stuttering-like disfluencies in everyday speech stem from covert repairs of errors of phonological encoding.

Published 2011 in the Journal of Communication Disorders
Brocklehurst, P. H., & Corley, M. (2011). Investigating the inner speech of people who stutter: Evidence for (and against) the Covert Repair Hypothesis. Journal of Communication Disorders, 44, 246-260. http://dx.doi.org/10.1016/j.jcomdis.2010.11.004

1 Introduction

1.1 Language production in people who stutter

The ICD9 (World Health Organization, 1977, p. 202), defined stuttering as "disorders in the rhythm of speech in which the individual knows precisely what he wishes to say but at the time is unable to say because of an involuntary repetition, prolongation, or cessation of a sound". Although this definition has been superseded, it reflects a continuing belief amongst many professionals who specialize in stuttering that people who stutter (PWS) do not have difficulty formulating in their minds (in inner speech) the words they want to say, but encounter difficulty when they attempt to express those words out loud. On face value, this would appear a reasonable assumption, as PWS do not generally report experiencing any difficulties formulating what they want to say. However, there is a substantial body of experimental evidence suggesting that PWS take longer to encode their utterances, and it seems likely that their language encoding abilities may indeed be overstretched by the time-pressures inherent in many everyday speaking situations. Less clear, however, is whether the inner speech formulated by PWS is qualitatively different from that of people who do not stutter (PNS).

In the present paper, we report an experiment designed to investigate the inner speech of PWS, determining whether they make more encoding errors than PNS. To clarify how this question arises and why it is relevant to an understanding of stuttering, we begin with an overview of evidence of language production impairment in PWS, and an outline of three psycholinguistic hypotheses that propose differing accounts of how such impairment may result in the disfluencies characteristic of stuttering.

1.2 Evidence of language encoding impairment in PWS

Many studies that have compared the spontaneous utterances of children who stutter (CWS) and age-matched controls have found evidence consistent with an interpretation of delayed/disordered phonological development or weaker encoding abilities in CWS (see Nippold, 1990; Nippold, 2001 for reviews). However, because spontaneous speech reflects the combined contributions of a variety of (linguistic and motor) factors, and because people who stutter have a tendency to avoid using sounds and words that they have experienced difficulty with in the past, it has been difficult to pinpoint exactly what it is that gives rise to such performance differences. In a similar way, the finding that PWS are often slower than age-matched controls at picture naming tasks (e.g. Bernstein Ratner, Newman, & Strekas, 2009; Newman & Bernstein Ratner, 2007), may reflect either slower lexical access or slower motor responses.

A number of studies have also shown that stuttering events appear to be linguistically determined. For example, stuttering is more likely to occur on more complex grammatical structures (Kadihanifi & Howell, 1992; Logan & Conture, 1997; Melnick & Conture, 2000; Ratner & Sih, 1987; Yaruss, 1999); on lower-frequency words (Anderson, 2007; Hubbard & Prins, 1994; Newman & Bernstein Ratner, 2007; Palen & Peterson, 1982); in association with (non-systematic) errors of phonological encoding (Yaruss & Conture, 1996); and on wrongly articulated word-initial consonant clusters (Wolk, Blomgren, & Smith, 2000). But since articulatory complexity tends to co-vary with these linguistic factors, the causal picture has remained unclear.

Somewhat more reliable evidence of slow or impaired language encoding in PWS comes from a series of priming and phoneme-monitoring studies, outlined below, all of which control for any possible influence of articulatory factors on response latencies through the use of repeated-measures paradigms in which the motor responses required of participants remain identical across all conditions.

Two sentence-structure priming studies have produced evidence of slow syntactic encoding. In these studies, children (Anderson & Conture, 2004) and adults (Tsiamtsiouris & Cairns, 2009) described action pictures immediately after listening to utterances containing syntactic structures that were either similar or different to those required for their own utterances. In both studies, speech-onset latencies of participants who stutter were longer in the absence of priming but the difference between the two groups reduced significantly when syntactically similar primes were used, suggesting that syntactic encoding is slower and perhaps less robust in PWS.

Priming studies investigating lexical access have produced less clear results. In one study, Pellowski and Conture (2005) found that when young CWS and age-matched controls performed a picture-naming task in which semantically related or unrelated words were auditorily presented just prior to picture presentation, mean speech-onset latencies of controls were shortened by semantic primes, whereas those of CWS increased. However, in a further study, comparing the effects of different types of lexical/semantic primes, Hartfield and Conture (2006) found that, in CWS, primes functionally related to target words interfered significantly less than those that were physically related. Hartfield and Conture (2006) concluded that these results, taken together, suggest that preschool CWS differ from controls in the speed and nature of lexical retrieval. Using a similar picture-naming paradigm, Hennessey Nang and Beilby (2008) compared the effects on adults who stutter (AWS) and age-matched controls of a variety of semantic and phonological primes, but failed to find any differences between the two participant groups: In both, semantically related primes resulted in longer naming-onset latencies. Nevertheless, Hennessey et al. (2008) suggested that group differences with respect to lexical-semantic encoding may exist, but only become apparent under conditions of high cognitive load (cf. Bosshardt, Ballmer, & De Nil, 2002; Weber-Fox, Spencer, Spruill III, & Smith, 2004).

With respect to phonology, although the phoneme priming studies that have been carried out have consistently found PWS' responses to be slower compared to controls' responses, the actual priming effects have been mixed. Melnick, Conture and Ohde (2003) found that both CWS and age-matched controls similarly exhibited shorter picture naming latencies when phonologically-related (CV) primes were played prior to picture presentation, suggesting that the two groups' phonological encoding abilities were comparable. However, more recently, Byrd, Conture, and Ohde (2007) found that, at six years of age, CWS continued to be responsive only to holistic primes, whereas six-year-old controls were responsive to segmental primes, suggesting that CWS are slower to adopt the use of segmental phonology, and instead continue to code words holistically. In AWS, The results of the phonological manipulation of the Hennessey et al. (2008) study mirrored those of Melnick et al. (2003). However, an (implicit) priming study by Wijnen and Boers (1994) found that AWS only benefited from primes containing both the initial consonant and vowel of the target word, whereas controls also benefitted from primes containing just the initial consonant, suggesting that AWS may have more difficulty encoding stress-bearing phonemes, although Burger and Wijnen's (1999) larger-scale rerun of same paradigm failed to reproduce these findings.

Despite the mixed results from phonological priming studies, two well controlled phoneme-monitoring studies (Sasisekaran & De Nil, 2006; Sasisekaran, De Nil, Smyth, & Johnson, 2006) found that, compared to PNS, PWS are significantly slower to identify phonemes in words formulated in their inner speech. Importantly, there were no differences between the two groups' performances with respect to auditory monitoring (of pure tones); picture naming; simple motor responses; or identifying phonemes when listening to tape recordings of the same words. These findings thus strongly suggest that AWS' slower responses on the phoneme monitoring task stemmed from impaired phonological encoding and not from any general monitoring impairment or slow motor responses.

Finally, compared to normally fluent speakers, both CWS (Anderson, Wagovich, & Hall, 2006; Hakim & Ratner, 2004) and AWS (Ludlow, Siren, & Zikira,

1997) have been found to be poorer at non-word repetition, although the extent to which poor performance on non-word repetition tasks may stem from impairment of phonological encoding remains somewhat unclear.

1.3 Language encoding impairment as a potential cause of stuttering

Although it remains unclear whether the language production abilities of PWS are qualitatively any different to those of PNS, the findings cited above generally suggest that, compared to PNS, PWS take longer to formulate their utterances and that this slowness may stem from slow phonological, lexical and/or syntactic, encoding. This conclusion is of particular interest insofar as it provides preliminary support for two psycholinguistic hypotheses: The Covert Repair Hypothesis (Kolk & Postma, 1997; Postma & Kolk, 1993), and the EXPLAN hypothesis (Howell & Au-Yeung, 2002), both of which posit a causal relationship between slow language encoding and stuttering.

According to the Covert Repair Hypothesis (CRH), slow language encoding increases the numbers of phonological encoding errors in speakers' speech plans. This is explained in terms of Dell's (1986) computer simulation of language production. In Dell's model, the activations of units (phrases, words and phonemes) gradually increase until they exceed those of any competing units. When overt speech is initiated, the units that happen to be most highly activated at that point of time are selected. Thus, the earlier the initiation of motor execution relative to the speed of encoding, the greater the chances that competing units will be selected in error.

The CRH attributes Stuttering-Like Disfluencies (SLDs) to covert repairs of language encoding errors. The notion of a covert repair is predicated on the idea that speech plans are frequently prepared in advance of their overt articulation and stored in an articulatory buffer for anything up to a few seconds before being articulated. During this time, the speaker can inspect these plans through an internal monitoring loop (roughly equivalent to inner speech) and cancel and reformulate them if necessary (e.g., Levelt, 1983; Levelt, 1989). If an error is perceived in this way and the speech plan is cancelled before the onset of overt articulation, a silent pause or “block” may ensue while the plan is reformulated. However, errors occurring later in the plan may not be noticed immediately. Thus situations may occur where overt articulation of the first phonemes, syllables or words of a plan may have already begun before the error is detected. In such cases, the speaker stops, retraces to a suitable point and starts again, the result being that, although the error itself is not articulated, the phoneme(s) or word(s) immediately preceding it will be repeated at least once and perhaps several times, depending on how many reformulations of the plan are needed before the correction is achieved. Repetition of continuants may occur without breaks in between, producing symptoms of prolongation rather than repetition. In this way, the CRH accounts for the three main types of stuttering-like disfluency: repetitions, prolongations and blocks.

Because covert repairs of syntactic and lexical errors are more likely to be associated with larger retraces (Nooteboom, 1980), Postma and Kolk (1993) proposed that the part-word repetitions characteristic of persistent developmental stuttering most likely stem from repairs of errors of phonological encoding. However, they did not rule out the possibility that covert syntactic or lexical error repairs may also result in the production of SLDs. Furthermore, as phonological encoding is effectively the end of the line with respect to production of the speech plan, slow syntactic or lexical encoding may impact upon the amount of time available for phonological encoding to be completed before motor execution begins. Hence, the CRH would predict that phonological symptoms are likely to be universally found in PWS, even when the primary site of impairment may be further upstream. [1]

---

[1] A study by Melnick and Conture (2000) that investigated possible links between syntactic encoding difficulty and phonological errors failed to find any such relationship, although it did find that, compared to CWS with age-appropriate phonology, CWS with concomitant phonological disorder were more likely to be disfluent when uttering syntactically complex sentences.

As the above review of the CRH indicates, this theory rests on two basic tenets: (1) the speech plans of PWS contain abnormally high numbers of phonological-encoding errors; and (2) the covert repair of such errors accounts for the characteristically high numbers of SLDs in their utterances. However, since Postma and Kolk's (1993) formulation of the CRH, no direct evidence has been produced to confirm these tenets.

The lack of evidence of any correlation between phonological encoding errors and SLDs has led to the development of two alternative psycholinguistic hypotheses of stuttering: the "Vicious Circle Hypothesis" (Vasić & Wijnen, 2005) and EXPLAN (Howell & Au-Yeung, 2002). The Vicious Circle Hypothesis retains the CRH's basic tenet that SLDs arise as the by-products of covert error repair. However, it proposes that, rather than making different numbers of inner-speech errors, the difference between PWS and normally fluent speakers is that PWS monitor their speech more vigilantly for errors and have a lower threshold for instigating repairs. PWS therefore perceive, and attempt to "repair", many minor sub-phonemic irregularities that normally-fluent speakers would not be concerned about. A further consequence of such hyper-vigilant monitoring is that disfluencies resulting from the error-repair process are themselves likely to be identified as errors, triggering further (unnecessary) reformulations of the speech-plan and leading to a "vicious circle".

The EXPLAN hypothesis also attempts to account for researchers' repeated failure to find firm evidence of abnormally high numbers of phonological encoding errors in the speech plans of PWS. However, according to EXPLAN, although SLDs arise as a direct result of slow language encoding, the majority are not the result of covert error repair. Rather, they occur at moments when the rate of speech planning has fallen below the rate of execution and the speaker has effectively run out of speech plan to articulate. At such moments, speakers tend to repeatedly execute whatever speech plan is already available to them until more becomes available for execution. Repetitions and prolongations (of phonemes, syllables and/or whole words) thus constitute an operantly-conditioned strategy (Howell & Sackin, 2001) that reduces silent pauses and thus help speakers maintain their conversation turns while completing the formulation of their utterances (cf. Blackmer & Mitton's 1991, autonomous restart mechanism). According to EXPLAN, the key factor that differentiates persistent stutterers from normally fluent speakers is that, when the rate of planning falls behind the rate of execution, whereas normally-fluent speakers habitually adopt a “stalling strategy” whereby they only repeat whole words that have already been formulated, persistent stutterers habitually adopt a maladaptive “advancing strategy” whereby they utter (and repeat) the incomplete fragment of the word currently being formulated. Unlike the stalling strategy, the advancing strategy leads to blocking and part-word repetitions and consequently a breakdown of rhythm.

Because evidence that the speech plans of PWS include more errors is crucial in distinguishing between the above theoretical accounts of stuttering, it is somewhat surprising that, to date, there has only been one directly relevant study published (Postma & Kolk, 1992). In this study, participants were instructed to repeat a series of CV and VC strings aloud, both with and without auditory masking, and to press a button each time they noticed themselves making an error. It was presumed that under conditions of auditory masking, participants would have had little choice but to rely on internal monitoring of the speech plan for error detection (although the possibility remains that some errors may have been detected through monitoring of kinesthetic or other forms of feedback). Although the PWS group reported a numerically higher proportion of phonemic errors than the controls, the difference was not statistically significant. However, the PWS group recited the strings more slowly, which may have enabled them to avoid errors which would have been manifest at a faster speech rate. Moreover, participants were not required to describe the errors that corresponded to button presses; effectively, the experimenter had to guess what errors had been perceived.

1.4 The current study

The current study was designed to examine whether the speech plans of PWS contain abnormally high numbers of errors of phonological encoding (the first tenet of the CRH).[2] In the study we used a modified version of a tonguetwister paradigm, originally developed by Oppenheim and Dell (2008) to investigate the effects of phoneme similarity on errors in inner speech. Participants recited tonguetwisters both overtly and in inner speech, and immediately reported any errors they perceived themselves making in either condition. Visual prompts were used to control speech rate, avoiding the possibility that participants would slow down to avoid making errors. In common with Postma and Kolk (1992), half of the participants were PWS, and half of all tonguetwisters were recited under conditions of auditory masking.

Because of the possibility that PWS differ from PNS with respect to monitoring vigilance (e.g. Lickley, Hartsuiker, Corley, Russell, & Nelson, 2005; Sherrard, 1975; Vasić & Wijnen, 2005), their overt recitations were also transcribed and coded by the experimenter. Thus for each group, the number of overt errors reported by participants could be compared to the number detected by the experimenter.

In the present experiment, we counted both onset errors and “word-order errors” (including word anticipations, perseverations and exchanges). This allowed us to investigate not only whether PWS have impaired phonological encoding, in line with Postma and Kolk’s (1993) original proposal, but also whether they have more difficulty producing sequences of words in the required order. We also took forward digitspan measurements for all participants, in order to control for differences in

---

[2] Fluency measures collected from participants also enabled us to conduct a (post-hoc) analysis, which constituted a preliminary test of the second tenet of the CRH: that stuttering-like disfluencies stem primarily from covert repairs of such errors. Details of this post-hoc analysis are provided in Section 4.2 of the Discussion.

participants' abilities to remember the tonguetwisters and to remember (and thus accurately self-report) their errors.

2 Method

2.1 Participants

Thirty-two people who stutter (eight male) and thirty-two controls (nine male) matched for age and education took part in the experiment. Participants were recruited through the University of Edinburgh student employment service and experimental subject-pool. Participants who stutter were additionally recruited through stuttering self-help groups, and some controls were recruited through an internet employment website. All participants were native speakers of English. PWS had a mean age of 38 (range 18 to 71); for PNS, the mean was 39 (range 18 to 68). Mean education level (on a scale where 1 corresponds to General Certificate of Secondary Education (GCSE) or equivalent, and 5 indicates a postgraduate degree) was 3.00 for PWS and 2.97 for PNS. Twelve participants from each of the groups were university students; four of the PWS and five of the PNS group were retired. There was a marginal difference between the two groups in forward digit span: mean digit span for PWS was 6.6, and for PNS, 7.2 (*t* 62 = 1.67; *p* =.061).

All participants completed Section 3a of the Overall Assessment of the Speaker's Experience of Stuttering (OASES; Yaruss & Quesal, 2006), in which respondents rate their current difficulty in communicating verbally on a five-point scale in each of 10 commonly occurring situations. These include, for example, talking with another person one to one, initiating conversations, speaking to strangers, and continuing to speak regardless of how your listener responds to you. Mean scores for the OASES section 3a (communication difficulty) were: PWS = 27.6; Controls = 19.7 (*t* (62) = 4.62 *p* <.001). In addition to these ten OASES 3a questions relating to general communication difficulty, participants additionally provided ratings of "fluency difficulty" in the same ten situations. Specifically, for each situation, they were asked to rate how difficult it is to speak fluently, without stuttering and without avoiding words. Mean scores for fluency difficulty ratings were: PWS = 31.9; Controls = 18.0 (*t* (62) = 8.65 *p* <.001)

For all participants who stutter, full SSI4 (Riley, 2009) stuttering severity measures were derived from video recordings of samples of their spontaneous conversation and reading out loud. Mean SSI4 score was 20.7; range 8 to 36. Control participants only completed the reading portion of the above assessment. Mean number of SLDs per hundred syllables on the reading task was: PWS 6.08 range 1.3 to 24; PNS 0.43 range 0 to 1.85; (*t* (62) = 4.15 *p* <.001).

Apart from stuttering, participants reported no speech, language, hearing or visual impairments that were likely to influence their results.

2.2 Materials

The experimental materials were identical to those used by Corley Brocklehurst, and Moat (in press), Experiment 1. They consisted of four-word tonguetwister sequences. The onsets of each tonguetwister sequence followed an ‘ABBA’ pattern to induce onset-phoneme substitution errors, e.g. *pink bid bit pick*. Sequences were generated automatically from a database of CVC(C) words with CELEX frequencies greater than 1 per million (Baayen, Piepenbrock, & Gulikers, 1995). Pronunciations were checked for ambiguity using the British English Example Pronunciation dictionary (BEEP: Robinson, 1997) and also by hand.

To enable additional analyses of phonemic and lexical influences on speech errors, we created four variants of each tonguetwister sequence, and divided these between four lists, each list containing 48 tonguetwisters (See Appendix 1). Each participant recited tonguetwisters from just one of these lists (i.e. each participant recited 48 tonguetwisters). The details of these phonemic and lexical manipulations are, however, not relevant to the current paper and are not described here.[3] Important for the analyses that are included in the current paper, is simply that every tonguetwister sequence was recited an equal number of times under identical conditions by (an equal number of) participants from both groups.

Within each list, half of the tonguetwisters were assigned to the auditory masking condition. Because auditory masking was blocked, four versions of each list were drawn up such that, in the experiment as a whole, all tonguetwisters appeared equally in masked and unmasked, and masking-first and masking-last conditions. Finally, two versions of each of the resultant 16 lists were created. In both versions, half of the tonguetwisters were marked for overt recitation and the other half for silent (inner-speech) recitation. Those that were marked for overt recitation in one version were marked for silent (inner speech) recitation in the other. This resulted in 32 lists of experimental items in a fully counterbalanced design.

Auditory masking was achieved using computer-generated pink noise, delivered through a set of Panasonic RP-HT225 stereo headphones. Participants' overt recitations of the tonguetwisters were captured on a Zoom H2 digital recorder and analyzed using Praat software (Boersma & Weenink, 2009).

2.3 Procedure.

The procedure was closely modeled on that of Oppenheim and Dell (2008) and Corley et al. (in press) with the exception that participants typed, rather than verbally reported, the details of any errors they perceived themselves to have made. This change was made to ensure that the ability to self-report errors would not be affected by stuttering.

Prior to beginning the experiment, participants underwent a computer-led tutorial and practice session, which included full instructions concerning the inner

---

[3] For a full explanation of the phonemic and lexical manipulations see Corley, Brocklehurst, and Moat (in press).

speech and overt speech procedures. In all conditions, participants were instructed to place highest priority on speaking in time to the (visual) metronome, not to worry about making mistakes, and simply to skip words they felt likely to stutter on. For the inner-speech recitations, to prevent them from attempting to mouth sequences silently, participants were instructed not to move their mouths or any muscles associated with speech and, if possible, to keep their mouths completely closed.

At the beginning of each masked block, participants were instructed to adjust the headphones to ensure that the loudness of the pink noise prevented them from hearing the sound of their own voice. It was emphasized that participants should speak the overt tonguetwisters as quietly as possible throughout the masked block. The experimenter observed participants throughout the experiment and reminded them, where necessary, to adhere to the above instructions.

Tonguetwister sequences were presented in a random order on a 17" computer monitor. For each sequence, participants underwent a *familiarization* phase followed by a *performance* phase. In the familiarization phase, the tonguetwister sequence appeared in the centre of the screen, above an icon prompting participants to speak overtly (a mouth). After three seconds, a series of four dashes appeared (one every second) below the tonguetwister, acting as a visual metronome for the repetition of the words in the sequence. In the masked condition, pink noise began as the first dash appeared, and lasted until the last of the four dashes disappeared. The dashes and mouth icon were then replaced by a single dot, which remained onscreen for an additional second before the mouth icon reappeared and the dash sequence started again. The dash sequence was repeated so that participants repeated each sequence aloud four times before the performance phase began. During familiarization, participants were not aware whether repetition of the sequence during the subsequent performance phase would be silent or out loud.

As soon as familiarization ended for each tonguetwister, it was moved to the top of the screen and the required speech modality (silent or overt) was indicated centre-screen by means of the mouth icon (as used in familiarization) or a face icon representing silent repetition. At the same time the words “press ENTER to continue" appeared below the icon. Pressing ENTER caused all text to disappear from the screen, leaving only the mouth or face icon visible. After 200 ms, a sequence of dashes appeared in the centre of the screen at a rate of one new dash every 500 ms, acting as a visual metronome for the (overt or silent) repetition of the four words of the tonguetwister sequence.[4] In the blocks with auditory masking, pink noise started to play over the participant's headphones as the first dash appeared. 500 ms after the appearance of the fourth dash, the dashes disappeared, the pink noise (if any) ended, and the tonguetwister sequence reappeared at the top of the screen, together with an instruction to “type any errors and then press ENTER to continue” at the bottom. If they perceived themselves to have made an error during a particular recitation, participants were instructed to type, as fully as possible, what they had actually said, for example “rag lap *rash* rap” (when they should have said “rag lap *lash* rap”; the /r/ in rash being an anticipation of the /r/ in rap). They were instructed to type one or more question-marks in the relevant places if they could not remember what they had said for a word or ‘X’s if they had completely omitted all or part of a word. Once errors, if any, had been reported, pressing ENTER started the next four-dash sequence. Each performance phase included four repetitions of the four dashes, before the familiarization phase for the next word sequence began.

In addition to participants' self-reports of their inner and overt speech errors, the experimenter independently identified and transcribed errors in the overt speech condition (this was done online and then double-checked from recordings). Recordings from a random sample of five PWS and five PNS were transcribed a

---

[4] This 500ms/word speech-rate has previously been found by Oppenheim and Dell (2008) to result in the production of a significant number of errors without leading to excessive co-articulation, blending, or elision of consonants.

second time by the experimenter prior to analysis to enable intra-rater reliability to be calculated.

2.4 Coding

Errors were coded into three categories: (1) onset errors; (2) word-order errors; and (3) other/ambiguous errors. Errors were only ascribed to the onset error or word-order error categories if they were not in any way ambiguous. Thus, for example, the onset error category only included instances where a ‘B’ word onset (i.e. the onset of words 2 or 3) was substituted by a (contextual or non-contextual) phoneme but the coda remained unchanged. Instances where the onset error resulted in production of one of the ‘A’ words (e.g. ‘dock dock notch dodge’ instead of ‘dock knock notch dodge’) were excluded, although such instances were rare because, in all but a few tonguetwisters, the codas of each of the four words differed.

Since the majority of errors occurred on words 2 and 3 (the ‘B’ words), counts of onset errors only included these words. This ensured that onset-exchange errors were only counted once. Counts of word-order errors included all four word positions, where one of the four words in the tonguetwister was either uttered twice, or exchanged positions with one of the other four words. Instances where the position change could potentially be accounted for as a result of a word omission (e.g. ‘rag lap rack xxxx’ instead of ‘rag lap lash rack’) and instances where the order-error could potentially have been a coda exchange (e.g. ‘rag lash lap rack’ instead of ‘rag lap lash rack’) were excluded from the counts. In cases where an error was followed by an overt self-repair, only the original error was coded for analysis.

2.5 Analyses.

Analyses were carried out using logistic mixed-effects regression modeling (Breslow & Clayton, 1993; DebRoy & Bates, 2004) using the lme4 package (Bates & Maechler, 2009) in R (R Development Core R Development Core Team, 2009). This approach allowed us to investigate the independent contributions of a variety of “predictor” variables (both naturally occurring and experimentally manipulated) to the (log) likelihood of making (a) phoneme-substitution errors, and (b) word-order errors. These two likelihoods thus constituted the two dependent variables.

For each dependent variable of interest we generated a base model which included an intercept, and random by-participant and by-item intercept variation. Because t-tests revealed a marginal difference between the two groups’ mean digitspan scores, digitspan was controlled for by including it first (as a covariate) in all analyses. We then proceeded to add predictors stepwise to each model under consideration. Predictors representing both main effects and relevant interactions were added in reverse order of theoretical importance; with covariates first and those of most relevance to the analysis being added last. Selection of models was based on two criteria. First, using χ^2 tests to compare model likelihood ratios, we assessed whether the fit of the model to the data was improved (as indicated by a significant decrease in the model likelihood ratio) by the addition of each predictor. With the exception of digitspan, predictors were retained only if the current (best) model was improved. Second, where two or more predictors each significantly improved the current model, we selected the model which had the smallest log-likelihood. Once predictors and their interactions had been exhaustively explored, the resulting model represented the ‘best fit’ to the data, being a model which could not be improved by the addition of further predictors.

Each model includes coefficients representing the intercept and any effects of predictors. Where models were selected, the Wald statistic, calculated from each estimated coefficient and its standard error, was used to determine whether the coefficients differed significantly from zero (see Agresti, 2002).

3 Results

In total, participants recited 12,288 four-word tonguetwisters (48 tonguetwisters, each repeated four times by 64 participants) and self-reported a total of 2201 errors of any type, of which 1230 were in overt speech and 971 in inner speech. Three sets of statistical analyses are reported which, together, address the question of whether or not the speech plans of PWS contain more errors of phonological encoding (the first tenet of the CRH). The first set of analyses reveals the factors that accounted for the distributions of participants' self-reported (internal and overt) errors. The second focuses on *overt* errors, and compares the numbers of participants' self-reports of their onset and word-order errors to those coded by the experimenter, to establish whether the groups differ with respect to the likelihood of detecting and reporting their own errors. The third focuses on the *accuracy* with which individual overt errors were reported by participants, i.e. the extent to which participants' self-reports exactly matched the experimenter's transcriptions from recordings.

For each of these analyses, data are presented separately for onset errors and for word-order errors.

3.1 Self reports (inner and overt speech combined)

In addition to random by-participant and by-item variation and digitspan, we included predictors of auditory masking (the presence or absence of pink noise while speaking); group membership (PWS or PNS); stuttering-like disfluencies per 100 syllables when reading aloud; self-ratings of difficulty speaking fluently; overtness (whether or not participants were speaking aloud); the overtness by group interaction; and the two overtness by disfluency interactions.

3.1.1 Onset errors

Out of a total of 281 self-reported onset errors, 160 were overt and 121 were in inner speech. Table 1 gives a breakdown of errors by experimental condition.

The best-fit model of self-reported onset errors included effects of digitspan, group and overtness (improvement due to adding overtness: χ^2 (1) = 13.87, *p* <.001).

No other predictors or their interactions significantly improved the model (all χ^2 (1) ≤ 2.65, $p \geq .103$.

Table 2 gives the coefficients of the model, and the probabilities that they differ from zero. After controlling for digitspan, compared to PNS, PWS were 2.37 (i.e. $e^{0.86}$) times more likely to report errors. Independent of group or digit span, participants were approximately 1.4 (i.e. $e^{0.36}$) times more likely to report errors in the overt condition.

3.1.2 Word-order errors

Out of a total of 218 self-reported word-order errors, 126 were overt and 92 were in inner speech. Table 1 gives a breakdown of word-order errors by experimental condition.

Insert table 1 here

In addition to the effects of digitspan, group, and overtness found for onset errors, there was an additional effect of masking (improvement due to adding masking: χ^2 (1) = 9.63, p =.002). As with onset errors, there were no significant interactions. No other predictors or interactions significantly improved the model (χ^2 (1) ≤ 2.76, $p \geq .097$). The model coefficients are given in Table 2, and show a similar pattern to that observed for onset errors: Compared to PNS, PWS were 3.05 (i.e. $e^{1.11}$) times more likely to report errors. Independent of group, masking or digit span, participants were more likely to report errors in the overt condition. Across both participant groups, masking caused an increase in the numbers of self-reported word-order errors, marginally less so with respect to errors in overt speech.

Insert table 2 here

3.2 Overt speech: Self reports vs. Experimenter reports

The group differences reported in sections 3.1.1 and 3.1.2 do not, in themselves, imply that the PWS group actually made more errors. It is also possible that PWS group members were simply better at detecting and self-reporting the errors they made, as suggested by the Vicious Circle Hypothesis (Vasić & Wijnen, 2005). To rule out this interpretation, we compared participants' self-reports of errors to the errors coded by the experimenter in all cases when speech was overt.

First, an intra-rater reliability check was carried out to ensure the experimenter's ratings were consistent across the two participant groups. Kappa values, based on re-transcriptions by the experimenter of recordings from a random sample of five PWS and five PNS (i.e. 15% of participants) were 0.81 for onset errors and 0.87 for word-order errors, indicating a high degree of consistency.

In addition to random by-participant and by-item intercept variation, the (mixed effects) regression models for these analyses included a third random variable with 65 levels, one for each of the individuals who identified errors: 64 participants, plus the experimenter. This variable thus controlled for differences in individuals' propensities to report errors. In addition to the fixed predictors discussed above, we also included a rater-type predictor, with 2 levels (self-rating or experimenter-rating).

The main interactions of interest were: group by rater-type (to clarify whether compared to PNS, PWS self-report a greater proportion of their overt errors); masking by rater-type (to clarify whether the proportion of errors self-reported by participants decreases when masking prevents them monitoring their own overt speech); and the three-way, group by rater-type by masking interaction (to clarify whether the extent of that decrease is different for PWS compared to PNS.)

3.2.1 Onset errors

The best fitting model of onset errors included the effects of group previously reported, showing that more errors were reported from the PWS group. This model also included an effect of rater (χ^2 (1) = 7.80, *p* =.005), indicating that, overall, the experimenter reported approximately 1.55 ($e^{0.44}$) times as many onset errors as did the participants (coefficients are given in Table 2). Importantly, there were no interaction effects (for all interactions χ^2 (1) ≤0. 20, p ≥.652). This implies that there was a consistent tendency across the two participant groups for participants to report fewer errors than coded by the experimenter.

3.2.2 Word-order errors

For word-order errors, the model including effects of digitspan and group was (marginally) improved by the addition of rater (χ^2 (1) =3.75, p =.053), indicating that, overall, the experimenter reported approximately 1.31 ($e^{0.27}$) times as many word-order errors as did the participants (coefficients are given in Table 2). Again, none of the interaction effects further improved the model (all χ^2 (1) ≤0.52, p ≥.471).

3.3 Accuracy of self reporting

Although the above comparisons (together) show that both PWS and PNS were around 0.7 times as likely to report errors as was the experimenter, they do not provide information on accuracy: There were, for example, some instances when a participant reported having made a different error from the one that was coded by the experimenter. We therefore recoded all instances of errors that had been identified by the experimenter and marked each one as *a 'match'* if it was identical to the participant's self-report, or a *'mismatch'* if it differed or was not reported by the participant. We also counted false alarms (i.e., instances where the participant self-reported an error which the experimenter coded as correct) but since there were only eight of these in total, they were not considered further.

We then performed a final pair of analyses on the likelihood of self-reporting an error that the experimenter had identified. Once again, separate analyses were conducted for onset errors and word-order errors. Numbers of matches and mismatches are shown in Table 3, and model coefficients are shown in Table 4.

Insert tables 3 and 4 approximately here

3.3.1 Likelihood of accurately identifying an error - Onset errors

After controlling for digitspan, the best fitting model included only *total errors* (the total number of experimenter-coded errors, of any type, in the tonguetwister) as a predictor, implying that, irrespective of which group participants belonged to, their reports of errors were less likely to identically match those of the experimenter on tonguetwisters that contained more overall errors. The model was not significantly improved by the addition of masking ($\chi^2 = 2.67$, $p = .102$), and, most importantly, the model was not improved by adding group membership as a predictor ($\chi^2 = 0.04$, $p = .837$) or any of the interaction terms that include group membership (all $\chi^2 \leq 2.17$, $p \geq .140$). This implies that, across experimental conditions, the two groups did not differ with respect to the likelihood that they would accurately report errors.

3.3.2 Likelihood of accurately identifying an error – Word-order errors

After controlling for digitspan, none of the predictors tested improved the fit of the model above that of the base model including an intercept (all $\chi^2 (1) \leq 1.56$, $p \geq .211$). The intercept was not reliably different from zero, reflecting the fact that overall the likelihood of participant-coded errors matching experimenter-coded errors did not significantly differ from around 50%.

4 Discussion

The primary aim of the current study was to test the first tenet of the CRH: that the speech plans of PWS contain abnormally high numbers of errors of phonological encoding. In Section 4.1 we argue that they do. In Section 4.2 we then present the results of a post-hoc analysis, the findings of which appear incompatible with the second tenet of the CRH: that stuttering-like disfluencies stem primarily from covert repairs of such errors. Although these post-hoc findings do not offer direct support for either the EXPLAN or Vicious Circle hypotheses, they are potentially compatible with both.

4.1 Group differences in speech error rates

The most important finding from the current experiment is that the PWS group self-reported significantly more onset and word-order errors than the control group, both in inner as well as in overt speech. Moreover, a significant proportion of the variance in the numbers of (both onset and word-order) errors made by participants was uniquely attributable to participants' group membership.

To establish whether the higher numbers of errors self-reported by the PWS group reflect higher numbers of errors actually made by that group, we investigated whether or not the two groups differed with respect to the vigilance with which they reported such errors. Our analyses revealed that, compared to experimenter-ratings, both participant groups under-reported their errors to a similar extent. When we investigated the accuracy with which participants reported individual errors, there was again no difference to be found between PWS and PNS. Taken together, these analyses show that the proportions of experimenter-coded errors self-reported by each group do not differ, a finding in line with previous studies that have compared the abilities of PWS and controls to detect phonemic errors in recorded speech (Postma & Kolk, 1992; Sasisekaran & De Nil, 2006).

Support for the assertion that this relationship holds for inner speech can be gleaned from the performances of the two groups when reporting their *overt* errors under conditions of auditory masking, where self-reports can be compared to experimenter ratings. This is because, under conditions of auditory masking, speakers are deprived of auditory feedback and thus forced to rely largely on internal monitoring (i.e., monitoring through the "inner loop") to detect their overt errors in the same way that they monitor for errors in inner speech. The absence of any significant masking by group interactions in the analyses of overt speech can be taken as further evidence that the two groups are vigilant to a similar degree when monitoring their inner speech.

Our findings therefore strongly suggest that the significantly greater number of onset-phoneme and word-order errors self-reported by the PWS group in inner speech reflects a greater number of (phoneme and word-order) errors actually occurring in the speech plans of that group. They thus support the first tenet of the CRH.

4.2 Covert error repair

The current experiment was designed to minimize the potential for covert error repair activity through instructions to participants to give priority to speaking in time to the (visual) metronome rather than to maintaining a high level of accuracy. It was therefore anticipated that the error patterns of the two participant groups would not reflect error repair activity to any significant extent. The finding that, within both participant groups, significantly more overt than inner-speech errors were self-reported was fully in line with our expectations in this respect, insofar as it suggests that the errors appearing in the original speech plan were not repaired prior to overt articulation. [5] Nevertheless, if covert error repair does play a significant role in causing disfluency in real-life speaking situations, we might have expected the participants with the highest scores on the measures of disfluency relating to real-life speaking situations to have made the most errors during the tonguetwister experiment. This was, however, not reflected in the results of our (mixed effects) regression analyses of speech errors. Whereas group membership (i.e. whether or not a participant was diagnosed as a PWS) was retained as a predictor in all of the best-fit regression models (indicating that group membership consistently predicted the likelihood of a participant making onset and word-order errors), neither fluency-difficulty self-ratings, nor percentage of SLDs in the reading task, nor any of the relevant interactions were retained in any models.

---

[5] It further suggests that a proportion of participants' overt errors were likely to have originated downstream from the speech-plan (in processes related to the generation of motor commands).

Because PWS are often relatively fluent when reading, it is possible that the reading task did not provide an adequate test of the relationship between speech-errors and stuttering. It is also possible that predictors relating to disfluency were rejected from the regression models because they shared too much variance with other predictors already present in the models. To further explore these possibilities, we therefore performed three post-hoc analyses using the data from only the PWS group. Specifically, we plotted PWS' inner and overt onset errors against: (a) their stuttering-like disfluencies per 100 syllables in the conversation task; (b) stuttering-like disfluencies per 100 syllables in the reading task; and (c) their self-ratings of difficulty speaking fluently (see Figures 1-3). Only two significant correlations were revealed in these plots: PWS' stuttering-like disfluencies in conversational speech were correlated to their (self-reported) overt onset errors (*r=0.44, p*=.012), and to experimenter reported overt onset errors (*r*=0.43, *p*=.015). Similar patterns were also evident in the plots of PWS' overt onset errors and their self-ratings of difficulty speaking fluently (Figure 3); however two outliers prevented these correlations from reaching significance.

The lack of any correlation between participants' inner-speech onset-error rates and their disfluency rates suggests that covert repairs of phonological encoding errors do not account for anything more than a minor proportion of the instances of SLDs in their everyday speech. These findings, therefore, are not compatible with the second tenet of the CRH: that stuttering-like disfluencies in PWS are the result of covert repairs of large numbers of errors of phonological encoding. If stuttering does involve covert error-repair, the errors in question must either be so subtle that participants failed to report them, or they must originate downstream from the speech plan, and therefore only in inner speech when it accompanies overt articulation (cf. Max, Guenther, Gracco, Ghosh, & Wallace, 2004).

In the current study, it was not feasible to attempt to measure whether or not the timing of the PWS group's inner-speech tonguetwister recitations was more variable than that of the PNS group. We therefore cannot rule out the possibility that stuttering may stem from inappropriate covert repair of small pauses and other prosodic markers in PWS' speech plans as posited in the Vicious Circle Hypothesis (Vasić & Wijnen, 2005; see also Lickley et al., 2005). Neither can we rule out the possibility that PWS are hypervigilant towards such cues, although the experimental paradigm could potentially be adapted to explore these possibilities. However if, as Vasić and Wijnen (2005) propose, stuttering-like disfluencies in PWS are related to hyper-vigilance with respect to subtle prosodic cues, an explanation is needed for why PWS do not show a corresponding hyper-vigilance with respect to the phonological errors that they make in inner and overt speech, and why subtle prosodic cues or irregularities would lead to high levels of error-repair activity when gross phonological encoding errors apparently do not.

Our failure to find any correlation between the frequency of inner-speech errors and SLDs is compatible with the EXPLAN hypothesis, and it is certainly possible that the higher numbers of onset and word-order errors made by PWS were simply a side-effect of language encoding impairment but did not in themselves contribute to the production of SLDs. In keeping with EXPLAN, it is also possible that a proportion of the perseveratory errors made by participants may have resulted from an established habit of repeating readily available segments or words in order to keep going when experiencing encoding difficulties (cf. Howell & Sackin, 2001). However, verification of these possibilities would require further research.

4.3 The role of working memory

Because the original intention behind the current experiment was to compare error rates of PWS with those of matched controls, differences in participants' digitspan scores have been treated as a potential confounding variable. The main concerns were that participants with shorter digitspans may (a) find it more difficult to remember tonguetwisters and therefore make more errors, and (b) have more difficulty remembering their errors and therefore tend to under-report (or mis-report) them. Thus in all regression analyses, any variance attributable to differences in participants' digitspans was first co-varied out by entering digitspan as the first predictor. However, it has been proposed (e.g. Bajaj, 2007) that impairments of (various aspects of) working memory may be directly implicated in stuttering. In light of such proposals it is noteworthy that digitspan was itself a significant predictor of self-reported error-rates; and, in line with studies that have reported poor non-word repetition abilities in PWS (Anderson, et al., 2006; Hakim & Ratner, 2004; Ludlow, et al., 1997), our current findings suggest that working-memory limitations may play a role in the higher speech-error rates found in PWS. This is perhaps not surprising bearing in mind the close association between working memory and phonological encoding, (Acheson & MacDonald, 2009a, 2009b).

4.4 Caveats

4.4.1 Ecological validity of the study

A number of studies have found that signs of language impairment only become apparent in PWS under conditions of increased cognitive load (e.g. Bosshardt, et al., 2002; Weber-Fox, et al., 2004). In light of these findings, it would appear that the current experimental paradigm must have been sufficiently cognitively demanding to reveal differences between the two participant groups. Nevertheless, the tonguetwister paradigm did not require participants to compose their own utterances or to attend to prosodic encoding, and neither did it require them to monitor whether they were being understood by their conversation partners or to attend to conversation partners' responses. It may thus have been less cognitively demanding than many of the speaking situations commonly encountered by participants in real life. It thus remains possible that language encoding errors may occur substantially more frequently in PWS in the cognitively demanding speaking situations encountered in their everyday life than they did in the tonguetwister experiment.

4.4.2 Word omissions

Together with the instruction to speak in time to the (visual) metronome, participants were also instructed to skip any words they felt likely to stutter on. This was to ensure that their overt recitations were uttered fluently and at the correct speech rate. We presumed that stuttering only occurs in overt speech, so we did not expect this instruction to cause participants to skip words in the silent (inner-speech) condition. However, we cannot rule out the possibility that, in the PWS group, a proportion of words may have been omitted because they were associated with past experiences of stuttering. Had these words not been omitted, the inner-speech error rates of participants who stutter may have been even greater.

4.4.3 The effectiveness of auditory masking

It is possible that auditory masking may not have completely blocked auditory feedback, especially bone-conducted feedback. Moreover, as kinesthetic feedback continued to be available to participants in the masked condition, participants may have continued to use either or both of these sensory feedback channels to some extent for error detection. In future studies, more effective blocking of sensory feedback, perhaps through a combination of methods, including auditory masking and tissue vibration, may reveal that the accuracy of self-reporting of inner-speech errors (i.e. in the absence of auditory or kinesthetic feedback) is somewhat lower than our data suggest. Whatever the case, in the present study, the (non-significant) trend toward poorer self-reporting of errors under auditory masking conditions was evident to a similar extent across both participant groups, so there is no evidence that the two groups utilized whatever sensory feedback was available to them to differing degrees. Nevertheless, future studies could use more effective feedback-blocking methods to control for such possibilities.

4.4.4 The nature of inner speech

In the current study we have adopted a Leveltian (1982, 1989) perspective on inner speech, equating it with Levelt’s “inner loop” insofar as it provides a way of inspecting the state of speech-plans held in an articulatory buffer prior to their release for motor execution. However, it is possible that inner speech may be derived from input from a different source or even a number of sources, depending on which are available (see Postma, 2000, for a review). Thus for example, during the overt masked conditions of the current experiment, inner speech may have been derived, at least in part, from corollary discharge from motor commands. Whatever the case, until the sources that contribute to inner speech can be clarified further, some uncertainty will remain regarding the extent to which errors perceived in inner speech reflect errors that have arisen during formulation of the speech plan rather than during formulation of motor commands.

5 Conclusions

When constrained to speaking at a fixed speech-rate, adults who stutter make more onset and word-order errors than do age-matched, normally-fluent controls. Moreover, the majority of these errors have their origin in the speech plan, before the onset of articulation. These findings represent important evidence in line with the first tenet of the Covert Repair Hypothesis (Postma & Kolk, 1993).

However, the lack of any correlation between the frequency of participants' inner-speech errors and the frequency of their disfluencies in real-life speaking situations suggests that covert repair of errors of phonological encoding is unlikely to contribute significantly to the manifestation of SLDs in everyday speech. These findings are thus incompatible with the second tenet of the Covert Repair Hypothesis.

In line with the EXPLAN hypothesis, it is possible that the higher numbers of onset and word-order errors made by PWS are a side-effect of language encoding impairment but do not in themselves contribute to the production of SLDs. However, it also remains possible that, in PWS, stuttering-like disfluencies may stem from covert repair of timing or prosodic errors, as proposed by Vasić and Wijnen (2005), or covert repairs of errors that occur downstream from phonological encoding, during the formulation of motor commands. Further research is needed to explore these possibilities.

6 References

Acheson, D. J., & MacDonald, M. C. (2009a). Twisting tongues and memories: Explorations of the relationship between language production and verbal working memory. *Journal of Memory and Language, 60*(3), 329-350.

Acheson, D. J., & MacDonald, M. C. (2009b). Verbal Working Memory and Language Production: Common Approaches to the Serial Ordering of Verbal Information. *Psychological Bulletin, 135*(1), 50-68.

Agresti, A. (2002). *Categorical data analysis*: Wiley-Interscience.

Anderson, J. D. (2007). Phonological neighborhood and word frequency effects in the stuttered disfluencies of children who stutter. *Journal of Speech, Language and Hearing Research, 50*(1), 229.

Anderson, J. D., & Conture, E. G. (2004). Sentence-structure priming in young children who do and do not stutter. *J Speech Lang Hear Res, 47*(3), 552-571.

Anderson, J. D., Wagovich, S. A., & Hall, N. E. (2006). Nonword repetition skills in young children who do and do not stutter. *Journal of Fluency Disorders, 31*(3), 177-199.

Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (release 2)[cd-rom]. *Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor]*.

Bajaj, A. (2007). Working memory involvement in stuttering: Exploring the evidence and research implications. *Journal of Fluency Disorders, 32*(3), 218-238.

Bates, D., & Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375–32.

Bernstein Ratner, N., Newman, R., & Strekas, A. (2009). Effects of word frequency and phonological neighborhood characteristics on confrontation naming in children who stutter and normally fluent peers. [doi: DOI: 10.1016/j.jfludis.2009.09.005]. *Journal of Fluency Disorders, 34*(4), 225-241.

Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition, 39*(3), 173-194.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1. 07). [Computer software manual]. Retrieved 21st May 2009, 2009, from http://www.praat.org/

Bosshardt, H., Ballmer, W., & De Nil, L. (2002). Effects of category and rhyme decisions on sentence production. *Journal of Speech, Language, and Hearing Research, 45*(5), 844.

Breslow, N., & Clayton, D. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association, 88*(421), 9-25.

Burger, R., & Wijnen, F. (1999). Phonological encoding and word stress in stuttering and nonstuttering subjects. *Journal of Fluency Disorders, 24*(2), 91-106.

Byrd, C. T., Conture, E. G., & Ohde, R. N. (2007). Phonological priming in young children who stutter: holistic versus incremental processing. *Am J Speech Lang Pathol, 16*(1), 43-53.

Corley, M., Brocklehurst, P. H., & Moat, S. (in press). Error Biases in Inner and Overt Speech: Evidence from Tonguetwisters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

DebRoy, S., & Bates, D. M. (2004). Linear mixed models and penalized least squares. *Journal of Multivariate Analysis, 91*(1), 1-17.

Dell, G. (1986). A spreading activation theory of retrieval in language production. *Psychological Review, 93*, 283–321.

Hakim, H., & Ratner, N. (2004). Nonword repetition abilities of children who stutter: an exploratory study. *Journal of Fluency Disorders, 29*(3), 179-199.

Hartfield, K. N., & Conture, E. G. (2006). Effects of perceptual and conceptual similarity in lexical priming of young children who stutter: preliminary findings. *J Fluency Disord, 31*(4), 303-324.

Hennessey, N., Nang, C., & Beilby, J. (2008). Speeded verbal responding in adults who stutter: Are there deficits in linguistic encoding? *Journal of Fluency Disorders, 33*(3), 180-202.

Howell, P., & Au-Yeung, J. (2002). The EXPLAN theory of fluency control applied to the diagnosis of stuttering. *Clinical linguistics: Theory and applications in speech pathology and therapy*, 75–94.

Howell, P., & Sackin, S. (2001). Function word repetitions emerge when speakers are operantly conditioned to reduce frequency of silent pauses. *Journal of psycholinguistic research, 30*(5), 457-474.

Hubbard, C., & Prins, D. (1994). Word familiarity, syllabic stress pattern, and stuttering. *Journal of Speech and Hearing Research, 37*(3), 564.

Kadihanifi, K., & Howell, P. (1992). Syntactic Analysis of the Spontaneous Speech of Normally Fluent and Stuttering Children. *Journal of Fluency Disorders, 17*(3), 151-170.

Kolk, H., & Postma, A. (1997). Stuttering as a covert repair phenomenon. In R. F. Curlee & G. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (2 ed., pp. 182-203). Boston; MA: Allyn & Bacon.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*(1), 41-104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge MA The MIT Press.

Lickley, R., Hartsuiker, R., Corley, M., Russell, M., & Nelson, R. (2005). Judgment of disfluency in people who stutter and people who do not stutter: Results from magnitude estimation. *Language and Speech, 48*(3), 299.

Logan, K. J., & Conture, E. G. (1997). Selected temporal, grammatical, and phonological characteristics of conversational utterances produced by children who stutter. *J Speech Lang Hear Res, 40*(1), 107-120.

Ludlow, C., Siren, K., & Zikira, M. (1997). Speech production learning in adults with chronic developmental stuttering. *Speech production: Motor control, brain research and fluency disorders,* 221–230.

Max, L., Guenther, F., Gracco, V., Ghosh, S., & Wallace, M. (2004). Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. *Contemporary Issues in Communication Science and Disorders, 31*, 105–122.

Melnick, K., & Conture, E. (2000). Relationship of length and grammatical complexity to the systematic and nonsystematic speech errors and stuttering of children who stutter. *Journal of Fluency Disorders, 25*(1), 21-45.

Melnick, K., Conture, E., & Ohde, R. (2003). Phonological priming in picture naming of young children who stutter. *J Speech Lang Hear Res, 46*(6), 1428-1443.

Newman, R. S., & Bernstein Ratner, N. (2007). The role of selected lexical factors on confrontation naming accuracy, speed, and fluency in adults who do and do not stutter. *Journal of Speech Language and Hearing Research, 50*(1), 196-213.

Nippold, M. A. (1990). Concomitant Speech and Language Disorders in Stuttering Children - a Critique of the Literature - Reply. *Journal of Speech and Hearing Disorders, 55*(4), 803-804.

Nippold, M. A. (2001). Phonological disorders and stuttering in children: what is the frequency of co-occurrence? *Clinical Linguistics & Phonetics, 15*(3), 219-228.

Nooteboom, S. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand.* (pp. 87–95). New York: Academic Press.

Oppenheim, G., & Dell, G. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition, 106*(1), 528-537.

Palen, C., & Peterson, J. M. (1982). Word frequency and children's stuttering: The relationship to sentence structure. [doi: DOI: 10.1016/0094-730X(82)90038-9]. *Journal of Fluency Disorders, 7*(1), 55-62.

Pellowski, M. W., & Conture, E. G. (2005). Lexical priming in picture naming of young children who do and do not stutter. *J Speech Lang Hear Res, 48*(2), 278-294.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition, 77*(2), 97-132.

Postma, A., & Kolk, H. (1992). Error monitoring in people who stutter: evidence against auditory feedback defect theories. *J Speech Hear Res, 35*(5), 1024-1032.

Postma, A., & Kolk, H. (1993). The covert repair hypothesis: prearticulatory repair processes in normal and stuttered disfluencies. *J Speech Hear Res, 36*(3), 472-487.

R Development Core Team. (2009). R: A language and environment for statistical computing. *Foundation for Statistical Computing, Vienna, Austria*. Retrieved from http://www.R-project.org

Ratner, N., & Sih, C. (1987). Effects of gradual increases in sentence length and complexity on children's dysfluency. *Journal of Speech and Hearing Disorders, 52*(3), 278.

Riley, G. (2009). StutteringSeverity Instrument for Children and Adults. (SSI-4) (4th ed). Austin TX: Pro-Ed

Robinson, A. (1997). The British English Example Pronunciation Dictionary, v0. 1. Retrieved 4th July 2009, 2009, from ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/dictionaries/

Sasisekaran, J., & De Nil, L. F. (2006). Phoneme monitoring in silent naming and perception in adults who stutter. [Research Support, Non-U.S. Gov't]. *Journal of Fluency Disorders, 31*(4), 284-302.

Sasisekaran, J., De Nil, L. F., Smyth, R., & Johnson, C. (2006). Phonological encoding in the silent speech of persons who stutter. *J Fluency Disord, 31*(1), 1-21; quiz 19.

Sherrard, C. (1975). Stuttering as “False Alarm” Responding. *International Journal of Language & Communication Disorders, 10*(2), 83-91.

Tsiamtsiouris, J., & Cairns, H. S. (2009). Effects of Syntactic Complexity and Sentence-Structure Priming on Speech Initiation Time in Adults Who Stutter. *J Speech Lang Hear Res, 52*(6), 1623-1639.

Vasić, N., & Wijnen, F. (2005). Stuttering as a monitoring deficit. In R. J. Hartsuiker, Y. Bastiaanse, A. Postma & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 226–247). Hove, East Sussex Psychology Press.

Weber-Fox, C., Spencer, R., Spruill III, J., & Smith, A. (2004). Phonologic processing in adults who stutter: electrophysiological and behavioral evidence. *Journal of Speech, Language, and Hearing Research, 47*(6), 1244.

Wolk, L., Blomgren, M., & Smith, A. B. (2000). The frequency of simultaneous disfluency and phonological errors in children: A preliminary investigation. *Journal of Fluency Disorders, 25*(4), 269-281.

World Health Organization. (1977). *Manual of the international statistical classification of diseases, injuries, and causes of death* (Vol. 1): WHO Geneva.

Yaruss, J. (1999). Utterance length, syntactic complexity, and childhood stuttering. *Journal of Speech, Language, and Hearing Research, 42*(2), 329.

Yaruss, J., & Conture, E. (1996). Stuttering and phonological disorders in children: Examination of the covert repair hypothesis. *Journal of Speech and Hearing Research, 39*(2), 349-364.

Yaruss, J., & Quesal, R. (2006). Overall Assessment of the Speaker's Experience of Stuttering (OASES): Documenting multiple outcomes in stuttering treatment. *Journal of Fluency Disorders, 31*(2), 90-115.

Table 1: Raw numbers of onset and word-order errors coded by participants and by the experimenter

		Onset errors			Word-order errors		
		Unmasked	Masked	(Total)	Unmasked	Masked	(Total)
Self-reports (inner speech)	PWS	37	44	(81)	27	46	(73)
	PNS	17	23	(40)	5	14	(19)
Self-reports (overt speech)	PWS	57	60	(117)	36	50	(86)
	PNS	22	21	(43)	20	20	(40)
Independent rater (overt speech)	PWS	77	88	(165)	46	60	(106)
	PNS	32	37	(69)	23	27	(50)

Table 2: Logistic mixed effects analyses of factors influencing the likelihood of occurrence of onset and word-order errors. Model coefficients and probabilities are given for best-fitting models. All intercepts represent the error probabilities for PNS, unmasked conditions.

Predictor	**Value**	**Co-efficient**	**Std. Error**	***p*** **(coefficient = 0)**
Inner vs. overt speech, onset errors				
(Intercept)	unmasked, PNS	-4.36	0.61	<.001 ***
	inner speech			
Digitspan	+1	-0.16	0.08	.041 *
Group	PWS	0.86	0.22	<.001 ***
Overtness	overt	0.36	0.12	.004 **
Inner vs. overt speech, word-order errors				
(Intercept)	unmasked,	-4.52	0.71	<.001 ***
	PNS			
	Inner speech			
Digitspan	+1	-0.26	0.10	.006 **
Masking	masked	0.44	0.14	.002 **
Group	PWS	1.11	0.31	<.001 ***
Overtness	overt	0.39	0.14	.006 **
Overt speech (participants vs. independent rater), onset errors				
(Intercept)	unmasked, PNS	-4.15	0.67	<.001 ***
	self-rated			
Digitspan	+1	-0.18	0.09	. 036 *
Rater	independent	0.44	0.11	<.001 ***
Group	PWS	1.05	0.24	<.001 ***
Overt speech (participants vs. independent rater), word-order errors				
(Intercept)	unmasked, PNS	-4.67	0.88	<.001 ***

	self-rated			
Digitspan	+1	-0.19	0.12	.114
Rater	independent	0.27	0.12	.032 *
Group	PWS	1.03	0.37	.005 **

Table 3: Accuracy of participants' coding of onset and word-order errors. Percentages represent the percentage of experimenter reports that were also self-reported (in an identical manner) by participants.

		Matches (%)	Mismatches (%)	Missed (%)	Total errors	False-alarms
Onset errors						
masked	PWS	46 (52)	21 (24)	21 (24)	88	3
	PNS	18 (49)	7 (19)	12 (32)	37	5
unmasked	PWS	46 (60)	7 (9)	24 (31)	77	0
	PNS	19 (59)	5 (16)	8 (25)	32	0
Word-order errors						
masked	PWS	32 (53)	22 (37)	6 (10)	60	0
	PNS	16 (59)	8 (30)	3 (11)	46	0
unmasked	PWS	24 (52)	10 (22)	12 (26)	27	0
	PNS	14 (61)	6 (26)	3 (13)	23	0

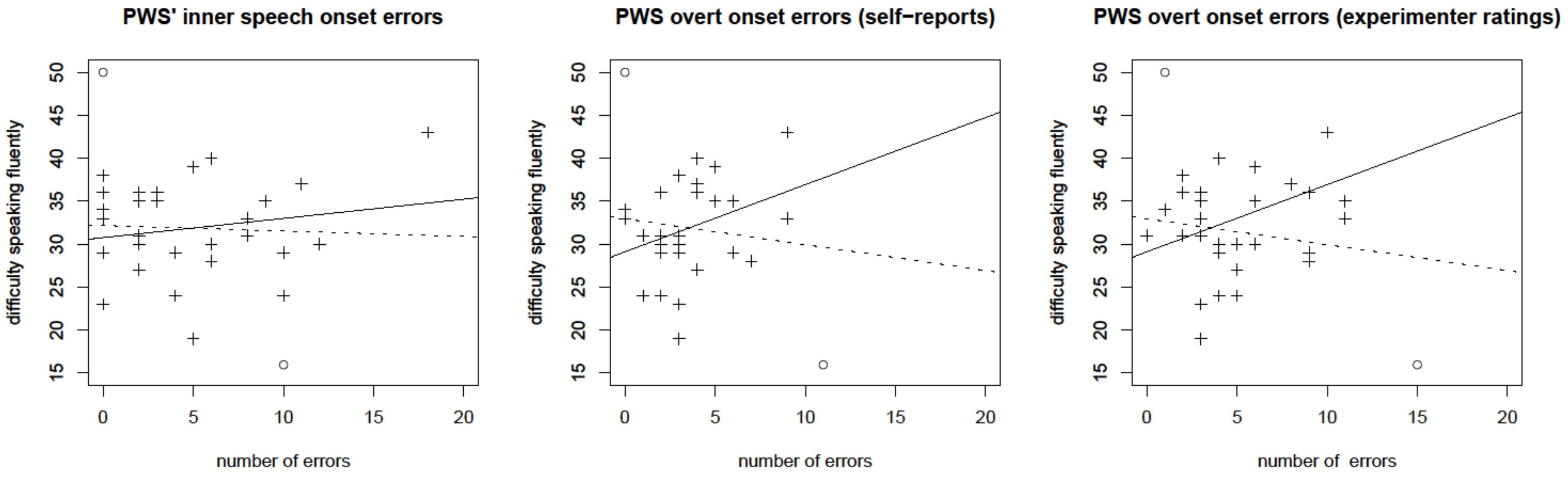
Table 4: Logistic mixed effects analyses of factors influencing the likelihood of participants accurately identifying onset and word-order errors. Model coefficients and probabilities are given for best-fitting models. All intercepts represent the error probabilities for PNS, unmasked conditions.

Predictor	**Value**	**Coefficient**	**Std. Error**	***p*** **(coefficient = 0)**
likelihood of accurately identifying an error (onset errors)				
(Intercept)	unmasked, PNS			
	inner speech	1.86	0.86	.030 *
digitspan	+1	-0.11	0.11	.330
Total errors in tongue-twister	+1	-0.21	0.08	.006 **
likelihood of accurately identifying an error (word-order errors)				
(Intercept)	unmasked, PNS	0..23		
	inner speech		0.19	.223

Figure 1: Scatterplots showing, for each participant who stutters, total number of onset errors (x axis) plotted against *stuttering-like disfluencies per 100 syllables*, as measured from the conversational speech task (y axis). From left to right, the three plots show raw numbers of onset errors in (1) inner speech; (2) overt speech (self-reports); and (3) overt speech (experimenter ratings).

Figure 2: Scatterplots showing, for each participant who stutters, total number of onset errors (x axis) plotted against *stuttering-like disfluencies per 100 syllables,* as measured from the reading task (y axis). From left to right, the three plots show raw numbers of onset errors in (1) inner speech; (2) overt speech (self-reports); and (3) overt speech (experimenter ratings).

Figure 3: Scatterplots showing, for each PWS, total number of onset errors (x axis) plotted against self ratings of difficulty speaking fluently in ten commonly occurring speaking situation (y axis).From left to right, the three plots show raw numbers of onset errors in (1) inner speech; (2) overt speech (self-reports); and (3) overt speech (experimenter ratings). The unbroken regression line depicts the relationship between the two variables when two outliers (marked with ○) are excluded.



Appendix A

Materials for Experiment

List 1				List 2				List 3				List 4			
fan	van	vat	fad	man	van	vat	mad	fan	van	valve	fad	man	van	valve	mad
pole	coast	cope	poke	soul	coast	cope	soak	pole	coast	comb	poke	soul	coast	comb	soak
till	kid	kin	tinge	bill	kid	kin	binge	till	kid	kiln	tinge	bill	kid	kiln	binge
seep	heath	heel	scene	keep	heath	heel	keen	seep	heath	heave	scene	keep	heath	heave	keen
rig	link	limb	rip	dig	link	limb	dip	rig	link	limp	rip	dig	link	limp	dip
pat	cap	catch	pad	bat	cap	catch	bad	pat	cap	cab	pad	bat	cap	cab	bad
busk	puff	puck	bunk	musk	puff	puck	monk	busk	puff	pub	bunk	musk	puff	pub	monk
cob	golf	gone	cot	yob	golf	gone	yacht	cob	golf	goth	cot	yob	golf	goth	yacht
finch	ship	shin	fill	pinch	ship	shin	pill	finch	ship	shift	fill	pinch	ship	shift	pill
meal	bead	beak	mean	weal	bead	beak	wean	meal	bead	beach	mean	weal	bead	beach	wean
dove	gulf	gull	dump	love	gulf	gull	lump	dove	gulf	gut	dump	love	gulf	gut	lump
wail	range	rake	waist	tale	range	rake	taste	wail	range	race	waist	tale	range	race	taste
pink	bid	bit	pick	kink	bid	bit	kick	pink	bid	bib	pick	kink	bid	bib	kick
come	tut	tub	cuff	hum	tut	tub	huff	come	tut	tuck	cuff	hum	tut	tuck	huff
conk	toss	top	cog	honk	toss	top	hog	conk	toss	tongs	cog	honk	toss	tongs	hog
reap	leap	leach	reef	beep	leap	leach	beef	reap	leap	leash	reef	beep	leap	leash	beef
dock	tod	tot	dodge	lock	tod	tot	lodge	dock	tod	Tom	dodge	lock	tod	Tom	lodge
peck	ketch	keg	pet	beck	ketch	keg	bet	peck	ketch	kelp	pet	beck	ketch	kelp	bet
gust	cusp	cut	gum	rust	cusp	cut	rum	gust	cusp	cup	gum	rust	cusp	cup	rum
face	vein	vale	feign	race	vein	vale	cane	face	vein	vague	feign	race	vein	vague	cane
pang	tank	tack	patch	hang	tank	tack	hatch	pang	tank	tap	patch	hang	tank	tap	hatch
hunk	thump	thug	hump	junk	thump	thug	jump	hunk	thump	thud	hump	junk	thump	thud	jump
rot	watt	wad	rob	not	watt	wad	knob	rot	watt	was	rob	not	watt	was	knob
tape	pain	pale	take	nape	pain	pale	knave	tape	pain	paid	take	nape	pain	paid	knave
tut	done	duck	tug	mutt	done	duck	mug	tut	done	dove	tug	mutt	done	dove	mug
dock	knock	knot	dodge	lock	knock	knot	lodge	dock	knock	notch	dodge	lock	knock	notch	lodge
sill	tick	tip	sick	chill	tick	tip	chick	sill	tick	tint	sick	chill	tick	tint	chick
deck	wreck	wren	dead	tech	wreck	wren	ted	deck	wreck	realm	dead	tech	wreck	realm	ted
run	duck	dub	rum	son	duck	dub	some	run	duck	dud	rum	son	duck	dud	some
wench	wreck	red	well	bench	wreck	red	bell	wench	wreck	rev	well	bench	wreck	rev	bell
kale	gauge	gape	cake	shale	gauge	gape	shake	kale	gauge	gait	cake	shale	gauge	gait	shake

bag	dad	dash	back	sag	dad	dash	sack	bag	dad	damp	back	sag	dad	damp	sack
mull	buff	buck	much	dull	buff	buck	Dutch	mull	buff	bulge	much	dull	buff	bulge	Dutch
roam	lone	lope	role	dome	lone	lope	dole	roam	lone	loaf	role	dome	lone	loaf	dole
wade	range	reign	wait	maid	range	reign	mate	wade	range	wraith	wait	maid	range	wraith	mate
sit	zing	zip	sick	knit	zing	zip	nick	sit	zing	zinc	sick	knit	zing	zinc	nick
puff	buff	bunch	punk	huff	buff	bunch	hunk	puff	buff	bulge	punk	huff	buff	bulge	hunk
rip	width	witch	rim	hip	width	witch	hymn	rip	width	wish	rim	hip	width	wish	hymn
dock	toss	tot	dosh	wok	toss	tot	wash	dock	toss	top	dosh	wok	toss	top	wash
delve	wreck	ref	dead	shelve	wreck	ref	shed	delve	wreck	realm	dead	shelve	wreck	realm	shed
wreck	wet	west	wren	peck	wet	west	pen	wreck	wet	wedge	wren	peck	wet	wedge	pen
fame	safe	sail	fade	maim	safe	sail	maid	fame	safe	sage	fade	maim	safe	sage	maid
bell	peg	pet	beck	knell	peg	pet	neck	bell	peg	pep	beck	knell	peg	pep	neck
pad	tank	tack	patch	mad	tank	tack	match	pad	tank	tab	patch	mad	tank	tab	match
teem	seep	seek	teach	beam	seep	seek	beach	teem	seep	siege	teach	beam	seep	siege	beach
hub	thump	thug	hush	rub	thump	thug	rush	hub	thump	thud	hush	rub	thump	thud	rush
jug	chuck	chump	just	lug	chuck	chump	must	jug	chuck	chub	just	lug	chuck	chub	must
rot	loft	lock	rob	not	loft	lock	knob	rot	loft	loll	rob	not	loft	loll	knob