Title:

Revisiting Bloodstein's Anticipatory Struggle Hypothesis from a psycholinguistic perspective: A variable release threshold hypothesis of stuttering

Manuscript accepted for publication in the Journal of Communication Disorders 06/04/2013

Authors:

Paul H. Brocklehurst[1,3]

Robin J Lickley[2]

Martin Corley[1]

[1] School of Philosophy, Psychology, and Language Sciences, University of Edinburgh. 7 George Square Edinburgh EH8 9JZ Scotland.

[2] Speech Science Research Centre, Queen Margaret University,

Edinburgh EH12 8TS Scotland.

[3] Stammering Self-Empowerment Programme, 27 Bridge Street, Macclesfield, Cheshire, SK11 6EG England.

Corresponding author:

Paul Brocklehurst

Department of Psychology

School of Philosophy Psychology and Language Sciences.

University of Edinburgh, 7 George Square, Edinburgh. EH8 9JZ. Scotland. UK.

Tel. 0044 798 615 3425 Fax 0044 131 650 3461

Email: P.H.Brocklehurst@ed-alumni.net

Revisiting Bloodstein's Anticipatory Struggle Hypothesis from a psycholinguistic perspective: A Variable Release Threshold Hypothesis of stuttering

Abstract

This paper reviews Bloodstein's (1975) Anticipatory Struggle Hypothesis of stuttering, identifies its weaknesses, and proposes modifications to bring it into line with recent advances in psycholinguistic theory. The review concludes that the Anticipatory Struggle Hypothesis provides a plausible explanation for the variation in the severity of stuttered disfluencies across speaking situations and conversation partners. However, it fails to explain the forms that stuttered disfluencies characteristically take or the subjective experience of loss of control that accompanies them. The paper then describes how the forms and subjective experiences of persistent stuttering can be accounted for by a threshold-based regulatory mechanism of the kind described in Howell's (2003) revision of the EXPLAN hypothesis. It then proposes that shortcomings of both the Anticipatory Struggle and EXPLAN hypotheses can be addressed by combining them together to create a 'Variable Release Threshold' hypothesis whereby the anticipation of upcoming difficulty leads to the setting of an excessively high threshold for the release of speech plans for motor execution. The paper also reconsiders the possibility that two stuttering subtypes exist: one related to formulation difficulty and other to difficulty initiating motor execution. It concludes that research findings that relate to the one may not necessarily apply to the other.

1. Introduction

The notion that anticipation plays a key role in the production of stuttered disfluencies has been a driving force behind several of the major hypotheses of stuttering that emerged in the twentieth century (Bloodstein, 1958, 1975, 1997; Johnson, 1942, 1959; Sheehan, 1953; Wischner, 1950). Of these, arguably the most fully developed is Bloodstein's (1975) 'Anticipatory Struggle Hypothesis', which posits that the anticipation of upcoming speech or communication failure causes people who stutter (PWS) to make adjustments to their way of speaking that result in the production of stuttered disfluencies. In this paper we describe the key points of Bloodstein's Anticipatory Struggle Hypothesis and review research that supports it and objections that have been raised to it. We then propose a way in which the hypothesis can be revised to bring it into line with recent advances in our understanding of the mechanisms underpinning speech and language production. This proposal, which we call the 'Variable Release Threshold' (VRT) hypothesis, entails a modification of the Anticipatory Struggle Hypothesis through the incorporation of a mechanism to account for the production of stuttered disfluencies, similar to that described by Howell (2003, 2004a) in his spreading activation revision of the EXPLAN hypothesis (Howell & Au-Yeung, 2002). A key feature of this VRT hypothesis is that the anticipation of imminent communication failure leads to an increase in the level of activation required before a speech plan can be released for overt articulation. We argue that, compared to Bloodstein's Anticipatory Struggle Hypothesis, the VRT hypothesis provides a more complete and parsimonious account of why stuttered disfluencies occur at the moments they do and why they tend to manifest in the specific variety of forms that they do. Finally, we consider how the VRT hypothesis may be of value, firstly with respect to its ability to provide an explanation for a number of observations about stuttering that researchers have found difficult to explain; and secondly with respect to its ability to clarify what we can realistically expect to achieve through therapy and what, if any, behavioral and/or environmental modifications may serve to ameliorate the condition.

2. Stuttering as an anticipatory struggle response

The term 'anticipatory struggle' is most commonly associated with Bloodstein, who used it in the 1950s to describe a broad category of hypotheses, all of which share the same underlying idea – that the anticipation of imminent speaking (or communication) difficulty interferes with the smooth running of the processes that underpin fluent speech (see Bloodstein & Bernstein Ratner, 2008, Chapter 2 for a review). Researchers have postulated a variety of mechanisms to account for how anticipation can lead to the production of stuttered and stuttering-like disfluencies, including an 'apprehensive, hypertonic avoidance' response (Johnson, Brown, Curtis, Edney, & Keaster, 1948); 'approach-avoidance conflict' (Sheehan, 1953); abnormal 'preparatory sets' (Van Riper, 1973), and 'tension and fragmentation' (Bloodstein, 1958, 1975). In the present paper we focus specifically on Bloodstein's (1975) Anticipatory Struggle Hypothesis – which posits that the primary symptoms of stuttering (repetitions, prolongations and blocks) result from a combination of inappropriate muscle tension and the fragmentation of planned utterances. Bloodstein (1975) proposed that tension and fragmentation are both essentially normal responses that speakers characteristically engage in when they wish to execute a complex motor activity and yet doubt that they will be successful:

If the persons who perform the activity believe it important to carry it off well… any serious reason they have to doubt they can do so may interfere with their performance. One form this interference takes is muscle tension. As they try too hard they may become too tense or tense in the wrong muscles, or in the wrong way. Another is fragmentation. They may take the activity apart and do it piece by piece. In particular they are likely to carry out a small bit of the beginning of the activity by itself and may reproduce this fragment repeatedly as long as the whole thing seems too difficult to do all at once." (Bloodstein, 1975, p. 4)

2.1.1. Origin of the anticipatory struggle response

Bloodstein (1975, p51) proposed that increased muscle tension and fragmentation of phrases or words into smaller more manageable units can be observed in the speech of most (perhaps all)

normally-developing young children when they attempt to produce utterances that they find motorically or linguistically demanding. He proposed that these normal responses are fundamentally the same as the tension and fragmentation responses which underlie stuttered disfluencies; the difference between the two being simply a matter of degree. Thus, in people who stutter, tension and fragmentation are so pronounced that their net effect is to interfere with the fluent production of speech to such an extent that they reduce rather than enhance its effectiveness. Bloodstein also proposed that the raised levels of tension and fragmentation associated with stuttering may stem from the persistent experience of speech or communication difficulty for any of a wide variety of reasons (cf. Johnson, 1942; for an alternative view). Such experiences are particularly likely to arise because of impaired or delayed development of linguistic skills and/or speech motor control, but they may also arise because of factors related to the listener or the speaking environment or a combination of all of these factors. Importantly, however, stuttering only starts to become persistent when these repeated experiences of speech or communication difficulty are sufficiently disruptive to cause the speaker to respond with tension and fragmentation to the mere anticipation that they *may* occur. Thus the Anticipatory Struggle Hypothesis attributes persistent stuttering to the belief that, in particular situations, particular sounds or words will be difficult to say or to communicate, and to the tensions and fragmentations that the speaker then produces in his attempts to compensate for or overcome the anticipated difficulty. By conceptualizing stuttering in this way, Bloodstein's Anticipatory Struggle Hypothesis allows for the possibility that, although the response (tension and fragmentation) may be similar in all people who stutter, the cues or impairments that provoke that response may differ widely from person to person. By taking the motives and perceptions of the speaker into account in this way, the Anticipatory Struggle Hypothesis provides a seemingly plausible explanation for a range of common observations about stuttering that purely impairment-based hypotheses find difficult to explain. In particular it provides an explanation for why PWS rarely stutter when speaking to themselves, or when they don't care what the listener thinks of them or what they say (Bloodstein, 1949, 1950);

and conversely, why they may find it so much more difficult to speak fluently to certain people, about certain topics in certain social situations (Bloodstein, 1949, 1950). It also potentially explains why some individuals who stutter perform as well as or even better than some normally-fluent speakers in tests of language and speech motor control ability.

In situations where the primary motivation is to communicate a message (as opposed to simply to speak), the Anticipatory Struggle Hypothesis posits that the responses of the listener can potentially be more important than the speaking abilities of the speaker with respect to their ability to precipitate disfluencies. By postulating that, in people who stutter, the anticipation of speech or communication failure precipitates a response that itself results in the failure that was anticipated, the hypothesis also provides an explanation for how stuttering may persist even in situations where any underlying language or speech impairment has resolved and where the listener and the speaking environment no longer pose any obstacles to communication.

2.1.2. Experimental evidence for Bloodstein's Anticipatory Struggle Hypothesis.

The key strength of the Anticipatory Struggle Hypothesis lies in its ability to provide a single, parsimonious explanation for a wide range of factors that influence of the severity of stuttering and the distribution of stuttered disfluencies, and also in its ability to provide a plausible account of how persistent stuttering may develop from speaking and communication difficulties experienced in early childhood. However, the hypothesis has proven difficult to test experimentally.

The main source of experimental evidence cited by Bloodstein (1975) in support of the Anticipatory Struggle Hypothesis is a series of 16 seminal experimental studies carried out by Johnson and his co-workers in the 1930s. In the first study of this series, Johnson and Knott (1937) found that, on successive readings of the same passage, PWS tend to stutter on the same words. This finding, which has since been repeatedly demonstrated experimentally, he termed the

'Consistency Effect'.[1] Using the same materials, Johnson and Millsapps (1937; Experiment 3) then found that if, on subsequent re-readings, previously stuttered words are blotted out and participants told to omit them, although the overall amount of stuttering was greatly diminished, participants now stuttered on other words, on which they had not previously stuttered, and a consistency effect became established on those words too. Moreover, a disproportionate number of these new stutterings occurred on words immediately adjacent to the blots. Johnson and Millsaps (1937) called this the 'Adjacency Effect' and interpreted it as an indication that stimuli representative of past instances of stuttering had the power to induce further instances of stuttering. Although this interpretation has been challenged (Wingate, 1986a), the findings of a further, related, experiment (Johnson, Larson, & Knott, 1937; Experiment 3) appear to reinforce it. In this related experiment, instead of blotting out the stuttered words, the experimenter simply marked them, and additionally marked five random, unstuttered words in the same way, and told the participant that each mark represented a previously stuttered word. On subsequent re-reading of the entire passage (including the marked words), stuttering on the five randomly marked words was three times more common than on five randomly selected (unmarked) control words, suggesting that cues that were perceived by the participant as representing past stuttering had the power to induce stuttering on whatever words were associated with them.

In a further investigation of the adjacency effect, Rappaport and Bloodstein (1971) performed a within-subject study with two conditions. Each participant read aloud a passage six times in the first condition and then a different passage six times in the second condition. After the third reading, the experimenter modified the script in one of two ways: In Condition (a) by blotting out the words that

1 Johnson & Knott (1937) also found an 'Adaptation Effect', whereby the overall likelihood of stuttering on a word reduces on subsequent iterations of it. This effect has since been well established and has been attributed to a number of possible causes (see Bloodstein & Bernstein Ratner, 2008, Chapter 11 for a detailed discussion).

had been stuttered in the previous three readings; or in Condition (b) by blotting out a similar number of words chosen at random. The order in which the two conditions were presented was reversed for half of the participants. For the participants who did the 'b' (random blot) condition first, there was no adjacency effect in either condition, whereas an adjacency effect was found *in both conditions* in the subjects who did condition 'a' first. These findings were interpreted as confirming that, once associated with stuttering, blots have the power to induce stuttering on adjacent words irrespective of where they occur in a text. Whereas, when they are not associated with stuttering, they have no such power.

Bloodstein (1975, p. 10) rejected the idea that the adjacency effect could be the result of classical conditioning. Such an account would require that the conditioned and unconditioned stimulus should be presented simultaneously or at least in close succession, which, in the Johnson and Millsapps (1937) experiment, was not the case. Rather, the blot obliterated the stimulus word and the two never appeared together. Bloodstein proposed instead that the perception of a relationship between the blots and past experiences of stuttering was cognitively mediated, and effectively constituted a belief. Moreover, the findings of the Johnson, Larson, and Knott (1937) study suggested that this belief could be falsely instilled by the experimenter, and the findings of the Rappaport and Bloodstein (1971) study suggested that once instilled, it tended to be self-sustaining.

A testable prediction of Bloodstein's (1975) Anticipatory Struggle Hypothesis is that the consistency effect should be stronger within speakers than between them. This is because the words on which people encounter difficulty in their everyday lives are likely to differ somewhat from individual to individual, and therefore the cues that have the power to evoke stuttering should also differ between individuals. The findings of a study by Hendel and Bloodstein (1973) supported this prediction: Across 17 AWS, on average, only 18 percent of words stuttered by one participant were stuttered by another, whereas on re-reading the passage, on average, each individual participant stuttered on 48 percent of previously stuttered words. Hendel and Bloodstein proposed that their

findings supported the hypothesis that the words stuttered were largely determined by individuals' personal past experiences of difficulty, and that such 'within-participant' factors played a more important role in determining which words would be stuttered than did factors associated with the contents of words themselves, such as word length, predictability, frequency etc. A similar study, by Stefankiewiez and Bloodstein (1974) found that when a four-week gap was interposed between readings, still 49% of words stuttered in the first reading were stuttered in the second, suggesting that 'within-participant' factors that elicit stuttering remain relatively stable over time.

2.1.3. The nature of the anticipated struggle

All of Johnson and associates' early studies investigating anticipation focused exclusively on the struggle to avoid stuttering and on the ability of cues that are evocative of past memories of stuttering to cause stuttering. The findings from these studies led Johnson to conclude that "expectation of stuttering is one of the psychological factors related to precipitation of the moment of stuttering" (Knott, Johnson, & Webster, 1937, p. 20), a view reiterated in the frequently cited statement of Johnson's that "stuttering is what you do trying not to 'stutter' again" (e.g. Johnson, 1972, p. 22). However, Bloodstein (1975) pointed out that Johnson's focus on anticipation and avoidance of *stuttering* reflected the relative ease with which this particular type of anticipation can be researched experimentally. He argued (pp 33-34) that to account for instances of stuttering in a more complete way, it must be possible that anticipation of upcoming struggle in a more general sense can act as a stimulus for stuttering, and he proposed that the tension and fragmentation that result in stuttering could equally be triggered by anticipation of any form of speech or communication difficulty. Importantly, Bloodstein's (1975) Anticipatory Struggle Hypothesis allows for the possibility that the anticipation of imminent struggle may be precipitated by the speaker's perceptions of the listener, including anticipation of listener miscomprehension, anticipation of a negative listener response, or indeed anticipation of any listener-related stimulus whatsoever that in the past has led the speaker to respond with tension and fragmentation. Although anticipation of

communication failure may often be closely tied to the anticipation of stuttering, this need not necessarily always be the case.

Reflecting this broader definition of anticipatory struggle, Bloodstein (1975) also identified two types of factor that interact in the development of stuttering in young children: (a) 'immediate' factors related to the child's abilities, such as delayed language or articulatory development; and (b) factors that create a more general atmosphere of communicative pressure, such as unrealistically high parental, societal, and self expectations. He also proposed that children who stutter may initially develop a generalized pervasive belief that speech is difficult and that such a belief may constitute "the germinal form from which more specific expectancies gradually develop" (p33).

Experimental studies to investigate Bloodstein's proposal that a wider range of experiences can lead to anticipatory struggle responses have been difficult to design due to problems with the operationalization of such value-laden concepts as 'communication failure' and 'negative listener responses'. Nevertheless, two studies have attempted to do so: An early experimental study by Hansen (1955), which investigated the influence of negative listener responses on the likelihood of producing stuttering-like disfluencies, and a recent study by Brocklehurst, Lickley and Corley (2012) which investigated the influence of word recognition failure.

In Hansen's experiment, participants who stutter performed a variety of reading and photograph-description tasks in front of an audience ranging from 12 to 25 people. The lighting was turned down so participants could not see the audience's faces. Positive or negative audience feedback was delivered to the speakers indirectly, by means of a series of green and red lights and corresponding counters, located on a table in front of the speaker. Each member of the audience was given a three-way switch to operate, with settings for positive, neutral and negative responses. The speaker and the audience were led to believe that the switches operated the lights, and the counters summed the three types of response for the duration of speaking. In reality, the colored-light activation and counter scores were regulated by the experimenter. Hansen found that,

although overall there was a general decrease in stuttering over the duration of the experiment, the rate of decrease was greater where favorable stimuli were delivered than where unfavorable stimuli were delivered. These trends became noticeable after a short time lag, and were most noticeable during spontaneous speech when it was easier for the speaker to focus on the audience responses.

In the Brocklehurst, et al. (2012) experiment, instead of speaking to an audience, participants were instructed to speak into what they believed was speech-recognition software on a computer, and received automatic online feedback indicating whether or not the words they had spoken had been correctly recognized. As in the Hansen experiment, the feedback was, in reality, determined by the experimenter, and bore no relationship to the accuracy, clarity or fluency with which participants spoke. Mimicking a speech-recognition system (rather than using human confederates) enabled the experiment to focus on the influence of the desire to communicate information (rather than the desire to avoid negative listener responses). To avoid any possibility that participants' performances may be affected by the fear of negative evaluation by potential listeners, participants were led to believe that they were not being recorded, that nobody was listening to them or able to hear them speak, and that the speech-recognition process was entirely automatic. Participants indicated before the first iteration of each new word whether they anticipated that they would stutter on it. They then spoke the word into the speech recognition system and immediately self-rated whether or not they had actually stuttered on it. After providing a self-rating, they received feedback from the software which indicated what word the software had 'recognized'. Participants produced four consecutive iterations of each word before moving on to the next word. Analysis of the results showed that three factors independently predicted the likelihood of stuttering on a word. Specifically, (a) 'iteration number' (Overall, the likelihood of self-reporting a word as 'stuttered' reduced across iterations); (b) participants' responses to the question "Do you think you will stutter on this word?" – which was posed prior to their first iteration of a word; and (c) the feedback participants received after providing self-ratings. Crucially, participants reported significantly less of a decrease in stuttering across iterations when feedback from previous iterations had indicated

incorrect recognition of the word than when it had indicated correct recognition[2]. This latter finding suggests that recent experiences of (apparent) failure to communicate a word increase the likelihood of stuttering on that word independently of the words lexical frequency, linguistic and articulatory difficulty, and the valence of listener responses.

2.1.4. Weaknesses of the Anticipatory Struggle Hypothesis

Despite its ability to provide an explanation for the loci of stuttered disfluencies and for many of the symptoms and experiences associated with persistent stuttering, Bloodstein's Anticipatory Struggle Hypothesis has been criticized on a number of counts.

Wingate (1986a, 1986b) questioned whether the Consistency and Adjacency effects are really related to anticipation. Instead, he proposed that both these effects are better accounted for in terms of linguistic stress[3]: participants simply stutter on the words they give more stress to. To illustrate this, he made a detailed analysis of recordings of PWS reading a passage aloud five times consecutively. The analysis focused on the patterns of disfluency produced by individual participants in addition to those of the group as a whole. This revealed that only a minority of individuals showed evidence of the Consistency effect, whereas all participants produced significant numbers of 'one time only' stutters, in all five readings. The consistency effect only emerged strongly when analyzing the stuttering loci of the group as a whole, reflecting the group's consistent use of stress on certain words in certain locations. When the same words appeared in unstressed positions in the passage, they were not stuttered to the same degree.

[2] Compared to the 'correct feedback' condition, in the 'incorrect feedback' condition the likelihood of stuttering increased by a factor of 1.5 with each subsequent iteration

[3] Wingate used the general term 'linguistic stress' to clarify that he was not referring to nervous stress (as in 'stress and anxiety'). Strictly speaking he was probably referring to prosodic stress rather than lexical stress.

On first reading, Wingate's (1986a, 1986b) linguistic stress hypothesis appears to provide an account of the Consistency and Adjacency effects that better fits the experimental data. However, contrary to what Wingate's argument implies, his and Bloodstein's hypotheses may not be incompatible with each other. Thus, for example, speakers tend to stress words that are critical for correct comprehension of a sentence. So if anticipation of communication failure can lead PWS to produce anticipatory struggle responses, then such responses would naturally be more likely to occur on stressed words. In this regard it is noteworthy that both linguistic stress and stutters are indeed more likely to occur on words that are critical to successful transmission of the intended meaning of the utterance (Cutler, 1984; Eisenson & Horowitz, 1945; Kaasin & Bjerkan, 1982). This issue is discussed in more detail below, in Section 5.1.3.

Packman Menzies and Onslow. (2000) have argued that the Anticipatory Struggle Hypothesis is incompatible with the substantial evidence of an underlying genetic predisposition and impaired fine motor skills in people who stutter. Their argument appears to be predicated on the (incorrect) assumption that Bloodstein's 1975 Hypothesis shares Johnson's (1942) notion that stuttering is entirely conditioned by listener responses and is completely unrelated to any underlying impairment (Packman & Attanasio, 2004). In fact, Bloodstein (1975) explicitly stated that the early experiences of struggle to speak or communicate may stem from, amongst other things, delayed speech, impaired articulation, aphasia, brain injury, cerebral palsy and mental deficiency , and "virtually anything at all that is calculated to shake children's faith in their ability to speak" (p40) . This statement clarifies a key difference between Johnson's and Bloodstein's hypotheses and implies that, unlike Johnson's, Bloodstein's Anticipatory Struggle Hypothesis is indeed compatible with recent genetic evidence and findings from brain imaging studies (see Bloodstein, 2000, for a fuller discussion).

Perkins (1997) made two criticisms of the Anticipatory Struggle Hypothesis that, in our opinion, are valid and need to be addressed. His first criticism relates to the notion, originally proposed by Johnson (1942), that factors that create a general atmosphere of 'communicative pressure' (such as

unrealistically high parental, societal, and self expectations) contribute to the onset of stuttering. His second criticism relates to Bloodstein's (1975) assertion that the primary symptoms of stuttering stem directly from tensions and fragmentations produced in response to the anticipation of upcoming difficulty.

With respect to the first of these two criticisms, Bloodstein (1975) identified two types of factors that interact in the development of stuttering in young children: (a) 'immediate' factors related to the child's abilities, such as delayed language or articulatory development; and (b) factors that create a more general atmosphere of 'communicative pressure', such as unrealistically high parental, societal, and self expectations. However, as intimated by Perkins (1997), over the last three decades, studies have consistently failed to demonstrate a link between parental interaction styles or communicative pressure and the onset of stuttering (e.g., Kelly & Conture, 1992; Kloth, Janssen, Kraaimaat, & Brutten, 1995, 1998; Weiss & Zebrowski, 1991). The failure of researchers to find evidence in support of a link between stuttering onset and parenting styles, or other communicative or environmental pressures has led to what appears to be a current consensus amongst researchers that such communicative pressures probably only influence the persistence of stuttering whereas its onset is more likely to be genetically determined or (perhaps occasionally) the result of trauma or injury (e.g., Alm & Risberg, 2007; Yairi & Ambrose, 2005). Consequently, researchers have largely rejected Bloodstein's account of the role communicative pressures in the ontogenesis of the disorder. However, it is possible that the failure to find evidence in support of a link between stuttering onset and communicative (and other environmental) pressures is because 'incipient' stuttering[4] and 'developed' stuttering are two distinct disorders (Bloodstein, 2001; see also

[4] The term "incipient stuttering" was used by Bloodstein to describe the transient, early form of childhood stuttering, which has traditionally been characterized as containing mainly easy and effortless repetitions of single-syllable words, syllables and sounds, as well as sound prolongations. More recently, studies of stuttering onset (Schwartz, Zebrowski, & Conture, 1990; Yairi, 1983; Yairi & Lewis, 1984) have determined that, contrary

Bloodstein, 2002, 2006) each with a separate onset. Thus, environmental factors may contribute significantly to the onset of developed stuttering, but may not play a significant role in the onset of incipient stuttering. This hypothesis will be discussed in detail in Subsection 5.1.1.

The second of Perkins' criticisms stems from his observation that the 'tension and fragmentation' response Bloodstein proposed to account for stuttered disfluencies appears to be incompatible with PWS' subjective experiences that a key element that differentiates stuttered from non-stuttered disfluencies is the feeling of a loss of control (Perkins, 1983; Perkins, Kent, & Curlee, 1991).

Perkins (1997) argued that although tension and fragmentation responses do occur, they are used in a volitional capacity and are not accompanied by the feelings of loss of control over the articulators that are a defining feature of stuttered disfluencies. Accordingly, he classified them as examples of *"nonstuttered disfluency that sounds like stuttering"* (Perkins, 1997, p223). In support of Perkins' objection, it is noteworthy that 'voluntary stuttering' (which involves the voluntary use of tension and fragmentation) has for many years been used as a way of reducing the feeling of loss of control (Van Riper, 1973). Furthermore, tension and fragmentation fail to adequately account for the specific forms (part and whole-word repetitions, prolongations of continuants, and 'blocks') that stuttered disfluencies characteristically take.

In the following section, we argue that the Anticipatory Struggle Hypothesis can be significantly strengthened by abandoning Bloodstein's notion that inappropriate tension and fragmentation constitute the primary mechanisms behind the production of stuttered disfluencies (although they may nevertheless play important roles in the production of secondary symptoms). In their place we

---

to traditional characterizations, the onset of stuttering can sometimes be sudden and severe, and that symptoms previously thought to be unique to persistent stuttering are also commonly found at or near the onset. (See Subsection 5.1.1)

propose that the primary symptoms of stuttering can be better accounted for in terms the malfunctioning of a release-threshold mechanism similar to that originally outlined by Howell's (2003) revision of the EXPLAN hypothesis. In order to present our argument in the clearest possible manner, we first outline the key points of the EXPLAN hypothesis and its release-threshold mechanism. We then consider how the anticipation of struggle may cause this mechanism to malfunction.

3. The EXPLAN hypothesis

The EXPLAN theory (Howell & Au-Yeung, 2002) is one of a number of psycholinguistic theories (e.g., Kolk & Postma, 1997; Postma & Kolk, 1993; Wingate, 1988) developed to account for the specific forms of stuttered and stuttering-like disfluencies, and for evidence linking such disfluencies with slow or impaired phonological encoding. According to EXPLAN, the majority of such disfluencies result from the failure of speech plans to achieve a sufficient degree of completeness to allow them be executed in a timely manner, and from the 'stalling' and 'advancing' compensatory behaviors that occur as a result. Stuttered (and stuttering-like) disfluencies occur when the speaker effectively runs out of adequately-formulated speech plan to utter. At such times, if the speaker attempts to continue, he will be unable to do so. However, he may maintain his conversation turn, by engaging in 'stalling' or 'advancing' behaviors that involve repeating or prolonging whatever sections of the speech plan are currently available until more plan becomes available (cf. Blackmer & Mitton, 1991).

Unlike other psycholinguistic theories, which attribute stuttering exclusively to slow or impaired speech planning, EXPLAN attributes stuttering to dyssynchrony between planning and motor execution. Although this dyssynchrony may stem from slow completion of the speech plan, it may also stem from the speaker adopting a speech-rate which is excessively fast. This latter scenario is most likely to occur, due to failure of the mechanism responsible for ensuring that planning and execution remain in step with one another. Failure of this mechanism may occur as a result of an

underlying weakness or pathology, possibly involving cerebellar dysfunction (Howell, 2004b), or from frequent experiences of speaking under time-pressure (See Section 3.1.2).

3.1.1. Error avoidance through the regulation of speech rate

Howell (2003, 2011) revised the EXPLAN hypothesis to bring it into line with Dell's (1986) spreading activation model of language formulation. Central to the revised (spreading activation) version of EXPLAN is the notion that speech plans cannot be released for overt articulation until the activation of their constituent units (phonemes, morphemes, words etc.) has exceeded a certain threshold (Howell, 2003, 2011; see Stemberger, 1985, for a similar proposal). Importantly, prior to execution, target units compete with other similar units for slots in the developing speech plan. As time progresses, the activation of target units increases beyond that of competing units. When execution is initiated, units with the highest levels of activation are selected for execution – provided their activation exceeds the threshold. Therefore, the likelihood of the wrong unit being selected for articulation decreases if more time is made available before execution is initiated.

Because target units generally reach higher levels of activation than competitors (Dell, 1986), the presence of a release threshold helps ensure that plans are relatively free from errors at the time they are executed. Thus, in addition to regulating speech rate and the timing of execution, the release-threshold mechanism also has the potential to fulfill a quality-control function, insofar as it helps ensure that only intended (i.e. target) units are executed.

3.1.2. Stalling and advancing behaviors

The rate at which activation of a specific unit builds up is dependent on how frequently it is used and how richly it is connected to other units. Thus, infrequently used units take longer to reach a level of activation high enough for them to be executed, and, as a result, they are less likely to be available when they are needed. When this happens, speakers may engage in 'stalling' behaviors whereby previous target units that are still sufficiently activated may be repeated or prolonged, to fill the gap until the desired target unit becomes sufficiently activated. Due to their high frequency,

function words are generally quicker to activate than content words, therefore situations often arise where function words are readily available for execution but content words are not. A common stalling scenario is that a function word may be repeated or prolonged until activation of the following content word reaches a high enough level to be executed.

Under normal speaking conditions, speakers will not attempt to execute a word until it is sufficiently activated to be executed in its entirety. However, under time pressure, speakers may engage in 'advancing' behaviors, whereby execution of the desired word's onset may be (repeatedly) initiated before the remainder of the word is sufficiently activated to allow its full execution. Depending on the extent to which activation has built up across the word's constituents, this may result in onset repetitions, prolongations, or 'blocks' (i.e., silent pauses stemming from the complete inability to articulate), which may continue until the activation levels of all of the word's constituents reach the required threshold (See Table 1). Together, these repetitions, prolongations, and blocks constitute the primary symptoms of stuttering.

Table 1. Stalling and Advancing disfluencies resulting from the unavailability of adequately formulated speech plan. As posited by the EXPLAN hypothesis (Howell & Au-Yeung, 2002; Howell, Au-Yeung, & Sackin, 1999)

Disfluency type	Behavior	Resulting disfluency
"Stallings" Common up to approximately 9 years of age (most evident in early childhood stuttering)	Re-issue whatever plans are already available until the intended target word becomes available.	Relaxed repetitions or prolongations of function word immediately preceding the target word.
"Advancings" Common from approximately 7 years of age onwards (especially prominent under time-pressure)	(1) Repeatedly attempt to execute target word in which the initial segment is incompletely formulated.	Tense pauses or 'blocks' prior to the target word.
	(2) Repeatedly attempt to execute target word in which non-initial segments are incompletely formulated.	Rapid initial-phoneme or syllable repetition, or prolongation of target word (accompanied by tension)

Howell and Au Yeung (2002) proposed that stalling behaviors (and the whole-word repetitions that result from them) are essentially benign. In contrast, advancing behaviors may lead to a pathological weakening of the speaker's ability to regulate his underlying speech rate and ensure that it is compatible with his formulation abilities. Howell, Au Yeung and Sackin (1999) equated the loss of this regulatory ability with the transformation of early childhood stuttering into persistent stuttering.

The EXPLAN hypothesis thus provides a plausible explanation for the variety of forms of disfluencies that characterize stuttering and also for the stuttering-like disfluencies that occur in normally-fluent speakers. Insofar as EXPLAN incorporates the notions of spreading activation and priming, it also explains the decreased likelihood of stuttering on subsequent iterations of previously spoken words (the 'Adaptation Effect'; Brutten & Dancer, 1980; Johnson & Knott, 1937), and some of well documented observations regarding the distribution and frequency of stuttered disfluencies, such as why they tend to occur on word onsets, and why the likelihood of stuttering occurring on a word is strongly influenced by its grammatical function (Bloodstein, 2006; Brown, 1937; Howell & Sackin, 2001), length (Brown & Moren, 1942; Danzger & Halpern, 1973; Soderberg, 1966; Wingate, 1967), position in the sentence (Brown, 1938; Quarrington, 1965), predictability (Quarrington, 1965; Schlesinger, Forte, Fried, & Melkman, 1965) and frequency (Brown, 1945; Newman & Ratner, 2007; Soderberg, 1966).

Compared to the Bloodstein's Anticipatory Struggle Hypothesis, the EXPLAN hypothesis is, however, less successful at accounting for a number of other important observations in relation to the distribution of stuttering events. In particular it fails to account for why adults who stutter frequently report being able to speak grammatically and motorically complex sentences completely fluently when they believe that nobody is listening to them. It also has difficulty accounting for why older children and adults frequently stutter on isolated, commonly occurring single words, and have particular difficulty uttering their names, why stuttering is influenced so strongly by the

characteristics of the listener and the overall dynamics of the speaking situation (Bloodstein, 1949, 1950; Bloodstein & Bernstein Ratner, 2008, Chapter 10), and why stuttered disfluencies sometimes continue to occur even at very slow speech rates.

In summary, the release-threshold mechanism, posited by the EXPLAN hypothesis, accounts for how speakers regulate their speech rate and minimize their speech errors. It is also successful at accounting for the forms that stuttered disfluencies take. However, it fails to fully account for why the rate of production of stuttered disfluencies is so dependent on the speaker's perceptions of the listener and his ability to hear, pay attention to, and understand what is said. In contrast, the Anticipatory Struggle Hypothesis fails to account for the forms that stuttered disfluencies take, but is fully successful at accounting for why the production of stuttered disfluencies is so dependent on the speaker's perceptions of the speaking situation and of the listener. In the following section we consider how Bloodstein's Anticipatory Struggle Hypothesis can be modified, by incorporating an EXPLAN-like release-threshold mechanism into it, so as to account fully for both the loci as well as the form of stuttered disfluencies.

4. Incorporating a release-threshold mechanism into an anticipatory struggle framework: The variable release threshold hypothesis

If a speaker perceives that his words are likely to be misheard, misunderstood or somehow fail to fulfill their intended function, irrespective of the actual cause of the anticipated failure, he is likely also to feel under pressure to in some way adjust his speaking style to rectify the situation. Thus, even if the anticipated failure is not in any way due to his own poor performance, he is still likely to perceive he can increase the chances of success by trying to speak as clearly and accurately as possible.

One relatively straightforward way of increasing the likelihood of being understood is to ensure that the speech plan is fully encoded and free from planning errors before initiating its overt execution. Because the likelihood of speech planning errors occurring decreases as the level of activation of a speech plan increases (Dell, 1986), it would make sense for the level of activation a plan has to achieve before it can be released for overt execution to increase in situations where the speaker perceives that his words are likely to be misheard, misunderstood, or fail to fulfill their intended function.

According to EXPLAN, stuttering results from the failure of speech plans to become activated beyond the threshold required for them to be released for execution. Thus, the variability in the severity of stuttering that occurs across speaking situations points to the likelihood that the level at which the release threshold is set varies from moment to moment: It rises at moments when the speaker perceives a need for a higher quality of speech – for whatever reason, and falls when speech quality is not considered important. If the release threshold is set at a relatively low level, speech can proceed at a faster rate, but is likely to contain more planning errors; whereas if the threshold is set at a high level, speech will contain fewer planning errors but will be released for execution more slowly. In PWS, stuttered disfluencies may occur when the release threshold rises to an abnormally high level in response to the perception of a need to speak more clearly and accurately. Such a response leads to longer onset times and, if the threshold rises too high, it may completely prevent words from being released at all – resulting in the experience of stuttering 'blocks' (see Figure 1).

INSERT FIGURE 1 HERE

Why might the release threshold rise too high? One possibility is that underlying impairments in a speaker's language or speech production systems mean that a rise in the release threshold does not result in the improved quality of speech that the speaker aspires to. In speakers whose language formulation systems are impaired or have yet to fully mature, target units may never achieve reliably higher levels of activation than competing units. As a result, many (categorical) speech errors will

continue to occur irrespective of how high the threshold rises[5]. In speakers whose language formulation systems are fully mature and functional, a rise in the release threshold will successfully increase the accuracy of encoding of the speech plan, thus reducing (categorical) lexical, syntactic, phonological and prosodic errors. However, it will have no effect on the quality of speech motor control, so the precision with which planned words are articulated will remain unchanged. If speech production is impaired at the level of speech motor control, then no matter how perfect the speech plan is, overt speech will still sound imprecise and may still be relatively difficult for listeners to understand. Unfortunately, although speakers may perceive a need for clearer, more accurate speech, they may not be sensitive to the level at which the problem has arisen (speech plan formulation or speech motor control). In the absences of such sensitivity, their response is likely to be stereotypical and sometimes inappropriate.

For the majority of speakers, the variable release threshold mechanism is likely to be effective as a general purpose quality control mechanism because a significant proportion of their communication failures probably stem from erroneous, inappropriate or incomplete encoding, which can be successfully rectified by a rise in the release threshold. As such, it may constitute the default mechanism for regulating speech accuracy. However, release-threshold regulation would be ineffective as a quality control mechanism for people whose communication failures result primarily from speech motor control impairment. In such situations, where a rise in the release threshold does

[5] A tonguetwister experiment by Brocklehurst and Corley (2011) revealed that, compared to normally-fluent speakers, adults who stutter (as a group) make significantly more onset substitution errors and word order errors, in both their overt and inner speech. However, there was significant overlap between the two groups, with some normally-fluent controls making more such errors than some PWS. This suggests that some (but not all) adults who stutter have an underlying mild impairment of speech planning. A similar finding of Group overlap in the extent of speech variability between AWS and controls in a study of articulatory variability by Kleinow & Smith (2000) suggest that some (but not all) AWS have a degree of impairment of their speech motor control.

not have the desired effect, speakers' ongoing (conscious or unconscious) attempts to speak more clearly and accurately, may cause the release threshold to continue to rise to higher and higher levels, reflecting their continued perception of a need for greater clarity and accuracy. A similar outcome may also result when the speaking environment is not conducive to successful communication of the intended message, perhaps because of excessive background noise or because of the listener's inability to hear or to understand what is said. And it is also possible that some PWS may simply harbor unrealistically high expectations regarding how 'perfect' their speech has to be.

5. Explanatory power of the VRT Hypothesis

In this section of the paper, we demonstrate how the VRT hypothesis has the potential to provide a single explanatory framework for a number of otherwise seemingly divergent observations relating to the onset, severity, and persistence of stuttering. Specifically, we focus on its potential to throw light on the following key issues: The relationship between early childhood stuttering and persistent stuttering; The nature of the primary and secondary symptoms of stuttering; The relationship between prosodic stress and stuttered disfluencies; The relationship between stuttering and inner speech; and the influence of auditory feedback, choral speaking and syllable-timed speech on stuttering.

5.1.1. The relationship between early childhood stuttering and persistent stuttering

Virtually all evidence that associates stuttered disfluencies with anticipatory struggle responses stems from experiments, observations and self reports from older children and adults who stutter. As far as we are aware, there is no reliable evidence to equate the onset of stuttering in early childhood with anticipatory struggle responses or the types of environmental or social pressures that Bloodstein suggested might lead to the development of such responses. On the contrary, over the

last three decades, studies that have investigated the influence of parenting styles and environmentally generated communicative pressures on the speech of very young children have repeatedly failed to find that these factors have any influence on the onset of stuttering (e.g., Kelly & Conture, 1992; Kloth et al., 1995, 1998; Weiss & Zebrowski, 1991). These negative findings suggest that the factors that influence the onset of stuttering in very young children differ from those that influence its persistence. Thus, the current general consensus among clinicians is that although parenting styles and environmentally generated communicative pressures may influence the persistence of stuttering they do not influence its onset.

As mentioned in Section 2.1.4, a possible alternative explanation for the failure of researchers to find evidence equating the onset of stuttering with environmental factors or parenting styles is that the 'incipient' stuttering of early childhood and the 'developed' stuttering that occurs in later childhood and beyond, are two distinct disorders (Bloodstein, 2001), and environmental factors and parenting styles only influence the second of these two disorders. This explanation would fit well with Bloodstein's (2001, 2002) proposal that incipient stuttering is associated with difficulties with syntactic formulation and the production of multi-word utterances. In support of this proposal, Bloodstein (2006) noted that episodes of incipient stuttering are often transient – coming and going repeatedly over a period of weeks or months, perhaps coinciding with critical moments in language development when the child is in the process of acquiring a new syntactic structure or rule. In contrast, the later form of stuttering, which may develop as a reaction to incipient stuttering and which is more likely to persist, is associated with difficulty with the initiation of individual sounds and words.

By adopting Bloodstein's (2001) notion of two distinct (but nevertheless closely related) etiologies of stuttering, the VRT hypothesis provides a new explanation for Howell, AuYeung and Sackin's (1999) finding that early childhood stuttering is characterized mainly by stalling symptoms, whereas stuttering in older children and adults is characterized mainly by advancing symptoms:

Namely, the symptoms of stuttering in early childhood mainly reflect stalling for extra formulation time while the child works out what words to say and what order in which to say them, whereas the symptoms of stuttering at a later age are more commonly associated with trying to 'push out' words that are already formulated: The speaker knows exactly which words to say, but cannot initiate their execution because their level of activation is still below the release threshold (see Figure 2).

INSERT FIGURE 2 HERE

According to the EXPLAN hypothesis (Howell, 2002; Howell & Au-Yeung, 2002), time-pressure is the key factor that leads speakers to adopt advancing behaviors in preference to stalling behaviors. In contrast, the VRT hypothesis posits that the key factor determining which of these two behaviors is adopted is whether or not syntactic formulation of the utterance has been completed. Although different, the EXPLAN and the VRT accounts of the etiology of stalling and advancing behaviors are not incompatible with each other, and both may be valid. Certainly the tendency to 'push' is likely to be increased in situations where there is time-pressure. However, whereas EXPLAN posits that advancings are primarily related to trying to speak too fast, the VRT hypothesis posits that they are primarily related to trying to articulate words that although adequately formulated, have nevertheless failed to achieve the release threshold.

So why might planning difficulties associated with early childhood stuttering constitute a predisposing factor to the development of persistent stuttering? The explanation that fits best with the VRT hypothesis is that the rise in the release threshold, which initially occurs as an appropriate anticipatory response to the desire to minimize planning errors when producing multi-word utterances, becomes generalized and starts to occur inappropriately. For example, it may start to occur as an anticipatory response to the desire to reduce speech motor errors – despite it having no effect on this type of error, or in response to the anticipation of listener miscomprehension or negative listener responses in situations where these responses are in actuality unrelated to the quality of the speaker's performance.

Bloodstein's account of how persistent stuttering develops out of early childhood stuttering was predicated on two closely related presuppositions about the nature of incipient stuttering: (a) that incipient stuttering forms a continuum with normal childhood disfluency; and (b) that, at the time of onset, the disfluencies of incipient stuttering are free from struggle, tension and other secondary symptoms. Although once widely accepted (Bluemel, 1932; Froeschels, 1943; Johnson, 1959), more recent findings from surveys and clinical studies conducted close to stuttering onset (e.g., Schwartz et al., 1990; Yairi, 1983; Yairi & Ambrose, 2005; Yairi & Lewis, 1984), have consistently revealed that early stuttering is in fact often associated with some degree of tension and force, and that the onset of stuttering in young children is often sudden and severe. These, and other related findings strongly suggest that early childhood stuttering is categorically different from normal childhood disfluency. They also suggest that, contrary to Bloodstein (2001), some very young CWS have difficulties with initiating the execution of single sounds and words. If this is the case, then it would appear that both types of stuttering (formulation-difficulty stuttering and execution-difficulty stuttering) can occur in early childhood.[6]

As mentioned previously, Bloodstein (2001) proposed that incipient stuttering predisposes to and may develop into persistent stuttering, a view reiterated both by Howell, Au Yeung, and Sackin (1999) and also by ourselves. However, although it is likely that this is often the case, clinical evidence suggests that execution-difficulty stuttering need not necessarily always be preceded by formulation-difficulty stuttering. Thus, for example Van Riper (1971), and Daly (1981) both identified subgroups of individuals (which Van Riper labeled 'Track 3') with late-onset developmental

[6] Because the stuttering-like disfluencies that appear in early childhood cannot be reliably equated solely with planning difficulties, it is useful to have suitable terms which clarify, where necessary, which etiology is intended. In the absence of more satisfactory terms, we adopt 'formulation-difficulty stuttering' for stuttering stemming from slow or impaired utterance formulation and 'execution-difficulty stuttering' for stuttering stemming from difficulty with the initiation of motor execution.

stuttering whose stuttering apparently began suddenly, often after a single traumatic event such as difficulty reading aloud in front of their school class. The existence of such cases points to the possibility that, contrary to what is generally believed, environmental pressures can indeed play a role in the onset of execution-difficulty stuttering. Whatever the case, if two distinct disorders do exist, then research that has failed to find a link between parenting styles or other environmental pressures and the onset of stuttering in early childhood cannot be validly cited as evidence that these factors do not play a role in the onset of late-onset stuttering (which is most likely to be of the execution-difficulty type). The relationship between environmental pressures and onset of this type of stuttering has not yet been investigated, and so remains unknown.

Finally, although 'persistent stuttering' almost invariably appears to be of the execution-difficulty type, this does not in any way imply that people do not ever recover from it. Indeed, it is possible (even likely) that recovery from execution-difficulty stuttering is the rule, rather than the exception, and that most recovery occurs in early childhood. If this is indeed the case, it would imply that although the presence of advancing symptoms in young children who stutter is a reliable indicator of the presence of execution-difficulty stuttering, it is probably not a strong or reliable predictor of persistence. This could account for why the Illinois studies (Yairi & Ambrose, 2005) failed to find any features (in terms of speech or secondary characteristics present around the time of onset) that served to distinguish children who will persist from those who will recover.

5.1.2. The primary and secondary symptoms of stuttering

Few researchers would question the traditional belief that repetitions, prolongations and blocks – the so-called 'primary' symptoms of stuttering – are symptoms that arise directly from the disorder, whereas concomitant movements and avoidance of words, people, and situations – the so-called 'secondary' symptoms – arise from the speaker's attempts to adapt to the primary symptoms.

An alternative view is that the experience of inability to move forward is the only truly primary symptom of stuttering, whereas prolongations, repetitions and visible, tense blocks are secondary

symptoms, reflecting the speaker's attempts to adapt to the inability to move forward. In recent years this latter view has tended to be disregarded, perhaps because of the lack of any objectively verifiable correlates to this experience (Moore & Perkins, 1990; Perkins, 1983, 1990). It nevertheless is the view that fits best with the EXPLAN and Variable Release Threshold hypotheses, insofar as repetitions and prolongations represent learned (potentially voluntary) responses that help the speaker maintain the attention of the listener and maintain his conversation turn until he is able to move forward (cf. Blackmer & Mitton, 1991; Howell & Au-Yeung, 2002), and the (more rapid) repetitions, prolongations and tense blocks characteristic of advancing behaviors represent the speaker's attempts to use force to push words out.

5.1.3. Regulation of prosodic stress

All other things being equal, speakers are likely to pay most attention to ensuring that they accurately select the units (syllables and words) that are most essential for listener comprehension and to ensuring that those units are appropriately emphasized. The VRT hypothesis therefore predicts that the release threshold for these essential units will be higher than for less essential units. In normally-fluent speakers, the relatively high level of activation of these essential units may result in a slight delay in their execution and an associated tendency for them to be executed with somewhat more force. The prominence resulting from this delay and increased force of execution may alert listeners to the relative importance of these units. In PWS, however, the higher release threshold of these essential units may render them impossible to execute. If this is the case, it may explain why PWS are more likely to produce stuttered disfluencies on stressed syllables and words (Wingate, 1986b), and why stuttering tends not to occur when speakers adopt speaking styles such as 'rhythmic speech' (e.g. Ingham et al., 2009) – which encourage speakers to focus more on the forward flow of an utterance than on the accuracy with which its specific units are encoded.

5.1.4. Persistent stuttering and the development of inner speech

Speakers normally have the ability to regulate the moment at which they initiate overt execution of a planned utterance. People who stutter, on the other hand, report that they are often unable to initiate the overt execution of their utterances, despite generally not having any difficulty producing them in inner speech. It thus appears that, for some reason or other, PWS have failed to develop the ability to regulate overt execution. Vygotsky (1986), proposed that learning to regulate the execution of speech plans is intimately connected with development of inner speech in early childhood. He proposed that young children initially speak their verbal thoughts out loud. Then, in response to social pressure not to constantly vocalize their verbal thoughts, they gradually develop the capacity to regulate the overt execution of speech plans, and to completely inhibit the overt execution of plans (i.e. thoughts) that are socially inappropriate or unhelpful. Vygotsky posited that when overt speech is inhibited in this way, the speaker nevertheless still 'hears' the plan in his inner speech and is able to use this inner speech for verbal reasoning and other cognitive operations.

From a Vygotskian perspective, the normal functioning of the VRT mechanism would enable regulation of the execution of verbal thoughts, and more specifically, regulation of whether and when verbal thought are expressed overtly and when they are restricted to inner speech. As such, control over the release threshold would constitute an important cognitive and social skill attained during childhood. If stuttered disfluencies stem from a failure of this mechanism to function in an optimal manner, we would expect them only to start occurring after a child begins to develop the capacity to inhibit overt speech in a socially appropriate way. Although such a causal connection may prove difficult to test experimentally, the observation that stuttering only begins a year or more after a child first starts uttering his first words (Bernstein Ratner, 1997; Yairi & Ambrose, 2005), and thus, presumably, after the child has started to become aware of the need to regulate execution, is fully in line with the Vygotskian account. Conversely, the observation that stuttering tends not to occur when the speaker forgets (or does not feel the need) to regulate his speech, such as in situations where there is no listener present, or during emotional outbursts is also in line with the Vygotskian account.

5.1.5. The influence of auditory feedback on stuttering

One of the unsolved mysteries of stuttering is why altered auditory feedback frequently (although not always) leads to a significant reduction in the severity of its symptoms. Researchers have proposed a number of possible mechanisms, the majority of which are predicated on the notion that unaltered auditory feedback alerts PWS to (real or perceived) errors or inadequacies in their speech, leading to inappropriate adjustments and the production of stuttered disfluencies (Civier, Tasko, & Guenther, 2010; Kalinowski, Armson, Stuart, & Gracco, 1993; Lincoln, Packman, & Onslow, 2006; Wingate, 1970). Speakers thus stop making such inappropriate adjustments if feedback is unavailable or distorted to the extent that the speaker no longer uses it (Van Riper, 1973). Alternative proposals are that altered auditory feedback results in a slower speech rate (e.g., Howell, 2007), or that it tricks the speaker into believing that he is speaking chorally, through the activation of mirror neurons (e.g., Kalinowski & Saltuklaroglu, 2003). In contrast, according to an anticipatory struggle account, altered auditory feedback simply removes some of the cues that might otherwise have alerted the speaker to similarities between his present speaking performance and previous performances in which he has struggled to speak or communicate in the past. In line with this anticipatory struggle account, the VRT hypothesis would predict that if cues contained in auditory feedback lead the speaker to anticipate a need to speak more accurately or clearly, they will result in an increase in the release threshold, which in turn will lead to stuttering. The reason delayed auditory feedback or frequency-shifted auditory feedback often lead to a reduction in stuttering may therefore be twofold: firstly, because such forms of feedback are not associated with past experiences of stuttering; and secondly, because the speaker knows that such forms of feedback are not providing him with useful information about the quality of his speech, so he does not rely upon them to make judgments about the adequacy of his speech. If altered auditory feedback does become associated with past experiences of stuttering, then it would lose its fluency-enhancing properties. The development of such associations may explain why altered auditory

feedback has been found to sometimes lose its effectiveness with continued use (Armson & Stuart, 1998).

The fact that approximately ten percent of PWS do not experience any increased fluency under altered auditory feedback conditions (Lincoln et al., 2006) suggests that not all PWS rely on auditory feedback as a means of determining the adequacy of their speech. For such individuals, other forms of feedback or monitoring channels may be more important. The VRT posits that, at any one time, there are a number of sources of cues that the speakers can draw on to inform him of the likelihood that their utterances will be good enough (e.g., proprioception, tactile feedback, efference copy, pre-articulatory error monitoring, conflict monitoring, monitoring of the listener and his responses), and the modality of the cues different speakers attend to will differ somewhat, depending upon their strengths and weaknesses and upon their past experiences.

5.1.6. Choral speaking and syllable-timed speech

For PWS, choral speaking and syllable-timed speech (e.g. speaking in time to a metronome) constitute the strongest fluency-enhancing forms of speech available (Andrews, Howie, Dozsa, & Guitar, 1982). Explanations that have been proposed to account for the effect are broadly similar to those proposed to account for the effects of altered auditory feedback. One property that both of these ways of speaking share in common is that they both force the speaker to give priority to maintaining the forward flow of speech (in order to keep up with the chorus or with the metronome beat). From the perspective of the Variable Release Threshold hypothesis, the fluency enhancing effect of these forms of speech can be explained by positing that, by forcing the speaker to give priority to the forward flow of speech they, by implication, also force him to give less priority to its accuracy and clarity. As a result of this change in priorities, the release threshold falls to a lower setting. In many ways, these fluency-enhancing ways of speaking place the speaker in a similar situation to a musician in an orchestra, whereby, if he plays a wrong or distorted note, or misses a note, he simply has to carry on as if nothing has happened.

6. The VRT hypothesis and the distal causes of stuttering

First and foremost, the VRT hypothesis proposes a single mechanism whereby the anticipation of upcoming difficulty speaking or communicating can lead to the production of stuttered disfluencies and other symptoms associated with moments of stuttering. With respect to the more distal causes of stuttering, the VRT hypothesis adopts a multifactorial stance insofar as it posits that any factors, or combination of factors (inherited, acquired or environmental) that cause speakers to anticipate difficulty speaking or communicating may predispose to stuttering. In the following two subsections, we consider how structural and functional abnormalities commonly found in PWS may cause the variable release threshold to malfunction; and how environmental pressures may cause it to malfunction. We also outline recent computational modeling studies that provides support for the role of a release threshold mechanism in the production of some types of stuttered disfluency.

6.1.1. Speaker-related factors

From the perspective of the VRT hypothesis, speaker-related factors that predispose to stuttering can be divided into two categories: those that do so because they impair the speaker's ability to plan or execute suitably well-formed utterances; and those that do so because they cause a speaker to be sensitive (or hypersensitive) to cues that alert him to the possibility that his speech performance is likely to be inadequate. Findings from neuroimaging and related studies provide converging evidence of three abnormalities in PWS that could impair their speech planning and execution abilities: (a) decreased myelination of white matter tracts underlying cortical areas responsible for speech planning and execution (Cykowski, Fox, Ingham, Ingham, & Robin, 2010; Watkins, Smith, Davis, & Howell, 2008); (b) excessive uptake of dopamine by cortical neurons (Alm, 2004; Wu et al., 1997); and (c) decreased myelination of cerebellar white matter tracts (Watkins et al., 2008).

All three of the above abnormalities could lead to malfunction of the variable release threshold mechanism. Specifically, *Insufficient myelination* of cortical white matter tracts may reduce the

efficiency with which target units in the speech plan become activated. Activation of target units may therefore fail to reach the release threshold in time for when they are needed for motor execution. Lower activation of target units may also lead to a reduced signal to noise ratio and an increased likelihood of competing units being selected in error. *Elevated dopamine levels* may result in target and competing units both becoming highly activated, again reducing the signal to noise ratio and increasing the likelihood of categorical speech errors. *Decreased myelination of cerebellar white matter tracts*, on the other hand, may result in excessive temporal and spatial variability in the motor execution of planned utterances, the net result being imprecise articulation and speech that sounds slurred. The speaker's perception of the poor quality of his articulation may then prompt an (inappropriate) increase in the release threshold (See Section 4).

In addition to playing a possible role in the impairment of speech production, elevated dopamine levels and cerebellar impairment may both also play roles in impairing speech perception insofar as they may cause speakers to become hypersensitive to cues that alert them to potential upcoming difficulty. Specifically, *elevated dopamine levels* may cause misinterpretation of auditory feedback[7], thus distorting speakers' perceptions of their performances; and *cerebellar impairment* may impair speakers' ability to utilize efference copy (Civier et al., 2010) or proprioceptive feedback (Loucks, De Nil, & Sasisekaran, 2007), thus causing them to rely excessively on auditory feedback instead.

Computational Modeling

Potential mechanisms by which incomplete myelination and excess dopamine could influence the speed with which speech plans become activated have recently been investigated by Civier

[7] The relationship between dopamine and auditory feedback sensitivity has yet to be explored in people who stutter. However, dopamine has been associated with distorted perception of feedback in a number of other dopamine-mediated conditions, such as schizophrenia (Frith, Blakemore, & Wolpert, 2000) and Parkinson's Disease (De Letter, Santens, & Borsel, 2005; Ho, Bradshaw, Iansek, & Alfredson, 1999).

(Civier, 2010; Civier, Bullock, Max, & Guenther, 2011) using GODIVA: a neurophysiologically plausible computational model of speech sequencing and planning (Bohland, Bullock, & Guenther, 2010) which incorporates a 'speech sound-map selection threshold' that is conceptually very similar to the release threshold of the VRT hypothesis. Simulations of excess dopamine led to delays to the moment when speech plans for initial syllables reached the selection threshold; and simulations of deficient white matter (in tracts underlying the ventral premotor cortex) led to delays to the moment when speech plans for non-initial syllables reached the selection threshold; with each abnormality resulting in a corresponding type of stuttering-like disfluency. Because the modules of the GODIVA model each correspond to specific brain structures, Civier et al. were also able to compare the activation levels of individual components within the two impaired versions of the model with blood-oxygenation level-dependent (BOLD) responses observed across those same brain regions in functional imaging data from PWS. The activation patterns produced by GODIVA correlated well with the corresponding imaging data.

In the above two studies, the threshold was not allowed to vary as in the VRT hypothesis. However, GODIVA is in principle capable of performing computational simulations in which the threshold is allowed to vary, as in the VRT hypothesis. So the effect of anticipation of communication failure on the release threshold should be amenable to such computer modeling.

A potential mechanism by which incomplete myelination of cerebellar white matter may lead to impaired speech motor control has been modeled (Civier et al., 2010) using the DIVA computational model. Modeling demonstrated how cerebellar impairment could lead to impaired feedforward control of motor execution and a corresponding excessive reliance on auditory feedback; the net result being slow vowel formant transitions. However, Civier et al. proposed that this would lead to the production of stuttered disfluencies through a process of 'motor resets' – a type of error repair. They did not consider the possibility that the slow transitions could impact upon the level at which the release threshold is set.

6.1.2. Listener-related factors

From the perspective of the VRT hypothesis, any listener-related or environmental factors that repeatedly cause the speaker to perceive a need to speak more clearly or accurately may contribute to the development of (execution-difficulty) stuttering in at-risk individuals (i.e., in individuals with any of the underlying weaknesses or impairments discussed in Section 6.1.1).

Repeated experiences of attempting to communicate with people who do not share the same first language may be particularly likely to give rise to such a perception, and this perception may constitute a contributory factor behind the higher incidence of stuttering that has been found in bilingual, compared to monolingual school children (Stern, 1948, cited in Bloodstein & Bernstein Ratner, 2008; Howell, Davis, & Williams, 2009; Travis, Johnson, & Shover, 1937)[8]. The VRT hypothesis predicts that stuttering risk should be decreased (in at-risk individuals) by a reduced focus on clarity and accuracy when trying to communicate in cross-linguistic speaking situations. This prediction should be amenable to testing, insofar as longitudinal studies may be able to demonstrate a lower incidence of stuttering in children taught to give priority, in cross-linguistic speaking situations, to the forward flow of speech compared to children taught to give priority to clarity and accuracy.

Similar perceptions of the need to speak more clearly and accurately may also result from repeated experiences of trying to communicate in noisy environments and/or with hearing-impaired listeners, and the VRT hypothesis predicts that children living in such environments may also be at a higher risk of stuttering. The relationship between stuttering and speaking in noisy environments is, however, likely to be a complex one, not least because different speakers may respond in different ways. Some may try to speak more clearly and accurately – a strategy which, as previously discussed, is likely to be counterproductive, especially in individuals with underlying motor control impairments

[8] Bloodstein and Bernstein Ratner (2008) and Yairi and Ambrose (2012) both point out that further studies are still needed to substantiate the assumption that Bilingualism is a risk factor for stuttering.

that predispose to stuttering. Others, however, may simply try to speak more loudly – a strategy which has sometimes been found to be fluency-enhancing, although not reliably so (Bloodstein & Bernstein Ratner, 2008, Chapter 11; Garber & Martin, 1977).

Exactly why speaking loudly might sometimes be fluency enhancing remains unclear. One possibility consistent with the VRT hypothesis is that, in some individuals, trying to speak with an appropriate level of loudness may cause the release threshold to rise (for the same reasons that trying to articulate accurately can cause it to rise). In individuals where this is the case, then abandoning any desire for appropriateness by purposefully adopting an *inappropriately* loud (or quiet) voice may allow the threshold to fall and result in increased fluency. It is also possible that some PWS, in common with people with other dopamine-mediated conditions, may have distorted perceptions of how loud their voices really are (See Section 6.1.1). Therefore, their attempts to speak with what they themselves believe to be an appropriate degree of loudness may in fact result in speech that sounds inappropriately quiet (or loud) to the listener and may thus fail to evoke the desired listener response. Where this is the case, the VRT predicts a therapeutic approach similar to that of Lee Silverman Voice Therapy (LSVT; Ramig, Pawlas, & Countryman, 1995) which aims to recalibrate speakers' perceptions of what constitutes an appropriate degree of loudness (by providing visual feedback in the form of decibel readings) may be beneficial, insofar as the availability of objectively reliable feedback confirming to the PWS that a satisfactory level of loudness has been achieved may help prevent such maladaptive rises in the release threshold.

7. Future Directions

7.1. Experimental research

7.1.1. Testing the role of anticipation

Central to the viability of experimental verification of the VRT hypothesis is the ability to manipulate experimentally, the extent to which subjects anticipate speech or communication failure. Brocklehurst et al.'s (2012) speech-recognition experiment demonstrates how subjects can be primed to anticipate word-recognition failure (or success) and how such priming can influence the likelihood of future stuttering on specific words. This paradigm could be extended so that participants' are primed to anticipate a range of different forms of communication failure, in a range of environmental settings. It should thus be possible to identify the extent of the range of different forms of anticipation that can influence the likelihood of stuttering, and it may be possible to quantify the specific contribution of each one of these to the likelihood of stuttering in specific individuals who stutter.

A further possible avenue for the exploration of the role of anticipation in stuttering is through studies that use electroencephalography to record event-related potentials (ERPs) produced in response to cues associated with stuttering. This type of approach, which is provides a high temporal resolution, has already been employed (with some success) to investigate pre-verbal error monitoring in PWS (Arnstein, Lakey, Compton, & Kleinow, 2011).

7.1.2. Verification of the VRT mechanism

As explained in Section 6.1.1, computer modeling using GODIVA (with a fixed release threshold) has demonstrated how failure of speech plans to attain an execution threshold in a timely manner can result in the production of stuttered (or stuttering-like) disfluencies (Civier, 2010; Civier et al., 2011). The researchers were then able to demonstrate how these simulated speech plan activation levels correlated with the blood-oxygenation level-dependent (BOLD) responses observed in brain regions associated with speech planning and execution in imaging studies of syllable and word production in PWS. If subsequent studies are able to provide converging evidence to confirm the location of the execution threshold mechanism, the key issue with respect to the VRT hypothesis is then to determine whether or not the release threshold mechanism is *variable*. To achieve this, it

should be possible to run simulations that vary the GODIVA release threshold in ways that correspond to experimental manipulations in speakers' levels of anticipation of upcoming difficulty with speech or communication. The results of these simulations can then be compared to brain-imaging and ERP results from real-life subjects, subject to the same variety of manipulations of anticipation of upcoming difficulty. This approach may also be able to distinguish the relative contributions to stuttering of a variable *release* threshold mechanism and a variable *repair* threshold mechanism (See Section 8.2.1). Ultimately, if the neurological correlates of the VRT mechanism can be identified, it should become possible to test the effect of various fluency enhancing techniques on the mechanism, and also to verify whether the decision to execute a planned utterance only in inner speech results in a corresponding increase or a decrease of the threshold. This, in turn, could provide valuable insights into the nature of inner speech (See Section 5.1.4).

A critical prediction of the VRT hypothesis is that no matter how slowly a PWS tries to speak, the level of fluency he is able to achieve will still be limited by the degree of accuracy that he aspires to. This contrasts with the EXPLAN account, which posits that there is a straightforward relationship between fluency and speech-rate, such that, provided the speaker can access the words he wants to speak, fluent speech can generally be achieved if he speaks sufficiently slowly. Specifically, the VRT hypothesis predicts that speakers will only be able to achieve a reliably high level of fluency to the extent that they are able to slow down and simultaneously reduce their self-expectations regarding accuracy. Indeed, according to the VRT, a reduction in effort towards accuracy is likely to be a more important factor than a reduction in speed in achieving an optimal level of fluency. In addition to having obvious clinical relevance, an experimental investigation of these two opposing predictions would help clarify this key point of difference between the VRT hypothesis and EXPLAN.

7.2. Clinical implications of the VRT hypothesis

Ultimately, the usefulness of the VRT hypothesis will be gauged in terms of its success in stimulating the formation of new and more effective approaches to treating stuttering. In the

current section we highlight some areas where the VRT hypothesis provides insights that may be relevant to therapy.

Insofar as the release threshold mechanism accounts for the production of stuttered disfluencies, it leads to two important questions, both of which have a direct relevance to therapy: (a) to what extent is the client who stutters trying to speak more accurately than he/she needs to? and (b) to what extent does he/she have the capacity to vary how accurately he/she tries to speak? With respect to the first of these two questions, the answer is likely to differ from individual to individual. Some PWS may have underlying language or speech production impairments that significantly limit their ability to produce accurate or clear speech under any conditions whatsoever. For such individuals, lowering the release threshold may result in speech that is impossible for listeners to understand. However, for the majority of PWS, it is more likely that the increased fluency achieved by relaxing their standards of accuracy will bring net benefits in many (although probably not all) speaking situations. For each individual client who stutters, the extent to which a relaxation of speaking standards does bring a net benefit could usefully be explored during therapy.

With respect to the second of these two questions, the VRT hypothesis predicts that, for people who stutter, key factors in achieving an improved level of communication effectiveness are the development of a more adaptive awareness of the relative importance of accuracy and fluency in specific speaking situations, and developing an awareness of how planning and motor control contribute to different aspects of the accuracy with which speech is produced. Clients may therefore benefit from forms of cognitive therapy that help them understand the antagonistic nature of fluency and accuracy, and, in particular, to understand that sometimes it may be possible to speak an utterance either fluently or accurately but not both fluently and accurately at the same time. They may also benefit from therapy that helps them to recognize the times when, due to factors related to the listener, or the environment, 'trying harder' to speak clearly and accurately is likely to be counter-productive. Similarly, clients with impaired speech motor control may benefit from

therapy that helps them understand their limitations with respect to the level of speech clarity they can hope to attain, and that explores ways of improving communicative effectiveness that do not precipitate a rise in the release threshold.

The VRT hypothesis predicts that a 'cost-efficient' way of maintaining fluency in real-life speaking situations may be through cultivating a willingness to reduce or abandon prosodic stress, especially on words that the speaker anticipates are likely to precipitate stuttering. Very few PWS are willing to employ traditional rhythmic speech strategies in real life situations, which is perhaps not surprising bearing in mind how unnatural they sound – and the likelihood that they draw the listeners' attention away from the message content. However, clinicians may be able help clients explore the extent to which a more limited reduction in prosodic stress can usefully be implemented in the various different speaking situations they encounter in everyday life.

The VRT hypothesis predicts that PWS may achieve an increase in fluency and an improvement in their overall communicative effectiveness through simply not to attempting to utter any utterance-constituent (phoneme, syllable or word) more than once. Thus clients could be instructed: "If a sound does not come out right first time, simply skip over it and continue on to the next sound (rather than going back and trying again)". By implication, this strategy, which has much in common with choral speech (See Section 5.1.6), favors fluency over accuracy and should lead to an overall reduction in the release threshold; especially if, having employed it, the PWS perceives that the listener has nevertheless satisfactorily understood what he has said. It is noteworthy that this strategy is diametrically opposed to Van Riper's (1973) strategy of 'Cancellation' – which, according to the VRT hypothesis, may result in a rise in the release threshold, and thus may be counterproductive.

8. Summary, caveats and conclusions

8.1. Summary

In this paper we have argued that, with respect to persistent stuttering, Bloodstein's (1975) Anticipatory Struggle Hypothesis provides the most satisfactory explanation for when and where stuttered disfluencies are likely to occur that is also fully compatible with wider evidence relating to the variety of hereditary and environmental factors that have been shown to influence the development of the disorder. We have, however, rejected the 'tension and fragmentation' account that Bloodstein (1975) proposed to explain the production of stuttered disfluencies, on the grounds that it is unable to account for the specific forms that stuttered disfluencies take, and because it fails to account for the subjective experience of loss of control. Instead, we have proposed that the forms that stuttered disfluencies take and the experience of loss of control that they entail can best be accounted for by a threshold mechanism similar to that proposed in the EXPLAN hypothesis; the normal function of which is to regulate the accuracy with which speech plans are encoded before they can be released for overt execution. To adequately account for both the loci and the forms of stuttered disfluencies, we therefore combined the key elements of Bloodstein's Anticipatory Struggle Hypothesis with those of Howell and Au Yeung's (2002) EXPLAN hypothesis to form a new hypothesis: the Variable Release Threshold hypothesis. The key feature of this VRT hypothesis is a release threshold mechanism that is responsive to the speaker's perception of a need to speak more clearly and accurately. We have proposed that the level of the release threshold is set differently for each syllable, depending on the speaker's perceptions of how important it is to ensure error-free production of that syllable. Syllables that are perceived as essential for the success of the speaking task are assigned higher release thresholds and are thus more difficult for PWS to execute.

The VRT hypothesis posits that stuttered disfluencies of persistent stuttering occur when the threshold for release of speech plans for overt articulation rises to a level that is so high that they cannot be released for execution at the moment they are required. Thus, beyond a certain point, efforts to increase clarity and accuracy of speech result in a maladaptive increase in disfluencies. This is especially likely to happen in speakers for whom the capacity for accurate articulation is limited by

some form of underlying speech motor control impairment, because a rise in the release threshold does not influence speech motor control and therefore cannot produce the desired increase in articulatory precision. It is possible that, once established, a tendency to set the release threshold too high may sometimes continue, even after any underlying impairment has resolved. It is also possible that recurrent experiences of communication difficulty stemming from environmental factors, such as listeners' poor comprehension abilities may lead some speakers to develop unrealistically high expectations of how accurately they need to speak.

8.2. Caveats

8.2.1. The role of error repair

Both EXPLAN and the VRT hypothesis are essentially 'error avoidance' hypotheses, in that they account for how PWS can reduce the likelihood of errors being encoded in the speech plan at the time of execution. In contrast, 'error repair' hypotheses posit that the production of stuttering-like disfluencies results from the process of repairing errors that are either encoded in the speech plan at the time of execution (Postma & Kolk, 1993; Vasić & Wijnen, 2005) or that arise during the process of motor execution (Civier et al., 2010).

Howell and AuYeung (2002) argued against Postma and Kolk's (1993) Covert Repair Hypothesis: that stuttering stems from the covert error repair of errors of phonological encoding, on the basis of lack of parsimony, and the relative rarity of such errors in the speech plan. More recently, two reviews of evidence for the Covert Repair Hypothesis (Brocklehurst, 2008; Vasić & Wijnen, 2005) have both concluded that the balance of experimental evidence does not provide convincing support for the hypothesis in its original form.

There is somewhat stronger support for error repair hypotheses that equate stuttering with repair of perceived timing errors (or delays), the frequency of which may be strongly influenced by the vigilance of monitoring (Vasić & Wijnen, 2005), or the accuracy of (and reliance upon) auditory

feedback (Civier et al., 2010). It is also possible that the two mechanisms: error avoidance and error repair, operate side by side – with stuttering being characterized by both an excessively high release threshold as well as an excessively low repair threshold; both thresholds being influenced (in opposite directions) by the anticipation of difficulty speaking or communicating.

On balance, we believe that, if these lower-level error repair mechanisms do play a role, it is likely to be a secondary one, insofar as they may account for some instances of repetition and prolongation. However, compared to the VRT hypothesis, they do not provide such parsimonious explanations for the subjective feeling of loss of control and the inability to initiate or move forward with articulation that PWS generally associate with stuttering. Nor do they appear to account for the full range of phenomena outlined in Section 5.

8.2.2. One release threshold or two?

The VRT hypothesis posits that there is only one release threshold for the execution of planned utterances and that, when execution is attempted, depending on whether or not the level of activation of the speech plan exceeds that threshold, the speaker will either (a) 'hear' the contents of the plan, internally, in inner speech; or (b) will produce it in overt speech. In other words, the experience of inner speech only occurs when a speech plan fails become sufficiently activated to exceed that threshold and fails to be executed overtly at the point of time when execution is attempted. On this account, inner speech will only be perceived when the release threshold rises to a higher level than the level reached by the speech plan. This account is in line with the Vygotskian theory outlined in Section 5.1.4.

An alternative possibility is that there are two release thresholds: one for inner speech and one for overt speech, and that the threshold for inner speech is set at a lower activation level than that for overt speech. If this is the case, then, effectively there are two executions: first a non-motor execution of the plan resulting in inner speech, and then (if it occurs at all) a motor execution of the plan resulting in overt speech. This account would fit better with Levelt's conceptualization (Levelt,

1989; Levelt, Roelofs, & Meyer, 1999) of inner speech as stemming from internal monitoring of the speech plan, prior to the initiation of motor execution.

Currently there is no experimental evidence that clearly favors either of these hypotheses over the other, although some preliminary support for the 'single threshold hypothesis' is provided by a recent picture-naming study, by Huettig and Hartsuiker (2009) who used eyetracking to compare the moment when speakers become aware of the contents of their speech plan when using only inner speech to name the pictures compared to when using overt speech. The findings suggested that speakers only perceive inner speech when overt speech is not produced. On the basis of parsimony, the VRT hypothesis has adopted the stance that there is a single release threshold and that inner speech stems from efference copy from failed execution of speech plans whose activation was below the level of the release threshold when execution was attempted. However, the locus from which inner speech is produced is currently an ongoing topic of debate in the field of psycholinguistic research (e.g., Corley, Brocklehurst, & Moat, 2011; Oppenheim & Dell, 2010; Pickering & Garrod, in press; Wheeldon & Levelt, 1995), and there is currently insufficient evidence to clarify which of the above options is most plausible.

8.3. Conclusions

In this paper we have described how Bloodstein's (1975) Anticipatory Struggle Hypothesis can be modified, through the incorporation of a variable release threshold mechanism, so that it can account not only for the loci, but also for the various forms that stuttered disfluencies take. We have also outlined how the resultant VRT hypothesis provides a new perspective on a number of further issues in relation to the disorder, including: the relationship between early childhood and persistent stuttering; the nature of the primary and secondary symptoms of stuttering; the relationship between stuttering and prosodic stress; the relationship between stuttering and inner speech; and the influence on stuttering of altered auditory feedback, choral speaking and syllable-timed speech.

As such, the VRT hypothesis has the potential to stimulate novel (and hopefully productive) approaches to research of these issues.

Insofar as the VRT hypothesis identifies stuttering as a condition that arises due to difficulty achieving an optimal balance between fluency and accuracy, the hypothesis also highlights the importance of establishing answers to two questions in relation to therapy: (a) To what extent is the client who stutters trying to speak more accurately than he/she needs to?, and (b) to what extent does he/she have the capacity to vary how accurately he/she tries to speak? The VRT hypothesis predicts that the most 'cost-efficient' ways of maintaining fluency in real-life speaking situations may be through cultivating a willingness to reduce prosodic stress on words that the speaker anticipates are likely to precipitate stuttering, and by continuing to move on to the next sound, regardless of how clearly or accurately the last sound or word was uttered.

9. Acknowledgement

The preparation of this article was supported by the UK Economic and Social Research Council (ESRC), grant number: ES/G01230X/1.

10. References

Alm, P. A. (2004). Stuttering and the basal ganglia circuits: a critical review of possible relations. *Journal of communication disorders, 37*(4), 325-369.

Alm, P. A., & Risberg, J. (2007). Stuttering in adults: The acoustic startle response, temperamental traits, and biological factors. *Journal of Communication Disorders, 40*(1), 1-41.

Andrews, G., Howie, P. M., Dozsa, M., & Guitar, B. E. (1982). Stuttering: Speech Pattern Characteristics Under Fluency-Inducing Conditions. *J Speech Hear Res, 25*(2), 208-216.

Armson, J., & Stuart, A. (1998). Effect of Extended Exposure to Frequency-Altered Feedback on Stuttering During Reading and Monologue. *J Speech Lang Hear Res, 41*(3), 479-490.

Arnstein, D., Lakey, B., Compton, R. J., & Kleinow, J. (2011). Preverbal error-monitoring in stutterers and fluent speakers. *Brain and Language, 116*(3), 105-115.

Bernstein Ratner, N. (1997). Stuttering: A psycholinguistic perspective. In R. Curlee & G. Siegel (Eds.), *Nature and treatment of stuttering: New directions (2nd ed.)* (pp. 99–127). Boston: Allyn & Bacon.

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

Bloodstein, O. (1949). Conditions under which stuttering is reduced or absent: A review of literature. *Journal of Speech and Hearing Disorders, 14*(4), 295-302.

Bloodstein, O. (1950). Hypothetical conditions under which stuttering is reduced or absent. *Journal of Speech and Hearing Disorders, 15*(2), 142.

Bloodstein, O. (1958). Stuttering as an anticipatory struggle reaction. In J. Eisenson (Ed.), *Stuttering: A symposium* (pp. 3-69). New York: Harper & Row.

Bloodstein, O. (1975). Stuttering as tension and fragmentation. In J. Eisenson (Ed.), *Stuttering: A second symposium* (pp. 1-96). New York: Harper & row.

Bloodstein, O. (1997). Stuttering as an anticipatory struggle reaction. In R. F. Curlee & G. M. Siegel (Eds.), *Nature and treatment of Stuttering: New Directions (2nd edition)* (pp. 167-181). Needham Heights, MA: Allyn Y Bacon.

Bloodstein, O. (2000). Genes Versus Cognitions in Stuttering: A Needless Dichotomy. *Am J Speech Lang Pathol, 9*(4), 358-359.

Bloodstein, O. (2001). Incipient and developed stuttering as two distinct disorders: Resolving a dilemma. *Journal of Fluency Disorders, 26*(1), 67-73.

Bloodstein, O. (2002). Early stuttering as a type of language difficulty. *Journal of Fluency Disorders, 27*(2), 163-167.

Bloodstein, O. (2006). Some empirical observations about early stuttering: A possible link to language development. *Journal of Communication Disorders, 39*(3), 185-191.

Bloodstein, O., & Bernstein Ratner, N. (2008). *A handbook on stuttering* (6th ed.). NY: Delmar.

Bluemel, C. (1932). Primary and secondary stuttering. *Quarterly Journal of Speech, 18*, 178–200.

Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural Representations and Mechanisms for the Performance of Simple Speech Sequences. [doi: 10.1162/jocn.2009.21306]. *Journal of Cognitive Neuroscience, 22*(7), 1504-1529.

Brocklehurst, P. H. (2008). A review of evidence for the covert repair hypothesis of stuttering. *Contemp Issues Commun Sci Disord, 35*, 25Á43.

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.

Brocklehurst, P. H., Lickley, R. J., & Corley, M. (2012). The influence of anticipation of word misrecognition on the likelihood of stuttering. [doi: 10.1016/j.jcomdis.2012.03.003]. *Journal of communication disorders, 45*(3), 147–160.

Brown, S. (1937). The influence of grammatical function on the incidence of stuttering. *Journal of Speech and Hearing Disorders, 2*(4), 207.

Brown, S. (1938). Stuttering with relation to word accent and word position. *The Journal of Abnormal and Social Psychology, 33*(1), 112-116.

Brown, S. (1945). The loci of stutterings in the speech sequence. *Journal of Speech and Hearing Disorders, 10*(3), 181.

Brown, S., & Moren, A. (1942). The frequency of stuttering in relation to word length during oral reading. *Journal of Speech and Hearing Disorders, 7*(2), 153-159.

Brutten, G., & Dancer, J. E. (1980). Stuttering adaptation under distributed and massed conditions. *Journal of Fluency Disorders, 5*(1), 1-10.

Civier, O. (2010). *Computational modeling of the neural substrates of stuttering and induced fluency.* Unpublished Doctoral dissertation, Boston.

Civier, O., Bullock, D., Max, L., & Guenther, F. H. (2011). *Dopamine Excess May Delay Selection of Syllabic Motor Programs: A Modeling Study of Stuttering.* Paper presented at the 17th International Congress of Phonetic Sciences, Hong Kong, China.

Civier, O., Tasko, S. M., & Guenther, F. H. (2010). Overreliance on auditory feedback may lead to sound/syllable repetitions: Simulations of stuttering and fluency-inducing conditions with a neural model of speech production. [doi: 10.1016/j.jfludis.2010.05.002]. *Journal of Fluency Disorders, 35*(3), 246-279.

Corley, M., Brocklehurst, P. H., & Moat, H. S. (2011). Error biases in inner and overt speech: Evidence from tongue twisters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 162-175.

Cutler, A. (1984). Stress and accent in language production and understanding. In Dafydd Gibbon & H. Richter (Eds.), *Intonation, accent and rhythm: Studies in discourse phonology* (Vol. 8, pp. 76-90). Berlin: Walter de Gruyter

Cykowski, M. D., Fox, P. T., Ingham, R. J., Ingham, J. C., & Robin, D. A. (2010). A study of the reproducibility and etiology of diffusion anisotropy differences in developmental stuttering: A potential role for impaired myelination. *NeuroImage, 52*(4), 1495-1504.

Daly, D. (1981). Differentiation of stuttering subgroups with Van Riper's developmental tracks: A preliminary study. *Journal of the National Student Speech and Hearing Association, 9*, 89-101.

Danzger, M., & Halpern, H. (1973). Relation of stuttering to word abstraction part of speech, word length, and word frequency. [doi: 10.2466/pms.1973.37.3.959]. *Perceptual and Motor Skills, 37*(3), 959-962.

De Letter, M., Santens, P., & Borsel, J. V. (2005). The effects of levodopa on word intelligibility in Parkinson's disease. *Journal of communication disorders, 38*(3), 187-196.

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

Eisenson, J., & Horowitz, E. (1945). The influence of propositionality on stuttering. *Journal of Speech and Hearing Disorders, 10*(3), 193.

Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. [doi:10.1016/S0165-0173(99)00052-1]. *Brain Research Reviews, 31*(2-3), 357-363.

Froeschels, E. (1943). Pathology and therapy of stuttering. *Nervous Child, 2*, 148-161.

Garber, S. F., & Martin, R. R. (1977). Effects of Noise and Increased Vocal Intensity on Stuttering. *J Speech Hear Res, 20*(2), 233-240.

Hansen, H. T. (1955). *The effect of a measured audience reaction on stuttering behavior patterns. (Unpublished doctoral dissertation).* University of Wisconsin, Madison.

Hendel, D., & Bloodstein, O. (1973). Consistency in relation to inter-subject congruity in the loci of stutterings. *Journal of Communication Disorders, 6*(1), 37-43.

Ho, A. K., Bradshaw, J. L., Iansek, R., & Alfredson, R. (1999). Speech volume regulation in Parkinson's disease: effects of implicit cues and explicit instructions. *Neuropsychologia, 37*(13), 1453-1460.

Howell, P. (2002). The EXPLAN theory of fluency control applied to the treatment of stuttering. In E. Fava (Ed.), *Clinical linguistics: theory and applications in speech pathology and therapy* (pp. 95-118). Amsterdam; Philadelphia; : J. Benjamins Pub. Co

Howell, P. (2003). Is a perceptual monitor needed to explain how speech errors are repaired? *Gothenburg Papers in Theoretical Linguistics, 90*, 31-34.

Howell, P. (2004a). Assessment of Some Contemporary Theories of Stuttering That Apply to Spontaneous Speech. *Contemporary Issues in Communication Science and Disorders, 31*, 123-140.

Howell, P. (2004b). Cerebellar Activity and Stuttering: Comments on Max and Yudman (2003). *J Speech Lang Hear Res, 47*(1), 101-104.

Howell, P. (2007). The Effects of Gated Speech on the Fluency of Speakers Who Stutter. *Folia Phoniatrica et Logopaedica, 59*(5), 250-255.

Howell, P. (2011). *Recovery from Stuttering*: Psychology Press/Taylor & Francis Group.

Howell, P., & Au-Yeung, J. (2002). The EXPLAN theory of fluency control applied to the diagnosis of stuttering. In E. Fava (Ed.), *Clinical linguistics: Theory and applications in speech pathology and therapy* (pp. 75–94). Amsterdam; Philadelphia: J. Bechamin's.

Howell, P., Au-Yeung, J., & Sackin, S. (1999). Exchange of stuttering from function words to content words with age. *Journal of Speech, Language, and Hearing Research, 42*(2), 345-354.

Howell, P., Davis, S., & Williams, R. (2009). The effects of bilingualism on stuttering during late childhood. *Archives of Disease in Childhood, 94*(1), 42-46.

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.

Huettig, F., & Hartsuiker, R. J. (2009). Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. [doi: 10.1080/01690960903046926]. *Language and Cognitive Processes, 25*(3), 347-374.

Ingham, R. J., Bothe, A. K., Jang, E., Yates, L., Cotton, J., & Seybold, I. (2009). Measurement of Speech Effort During Fluency-Inducing Conditions in Adults Who Do and Do Not Stutter. *J Speech Lang Hear Res, 52*(5), 1286-1301.

Johnson, W. (1942). A study of the onset and development of stuttering. *Journal of Speech and Hearing Disorders, 7*(3), 251.

Johnson, W. (1959). *The onset of stuttering*. Minneapolis MI: University of Minnesota press.

Johnson, W. (1972). Desirable Objectives and Procedures For an Adult Stutterer. In M. Fraser (Ed.), *To the Stutterer (Stuttering Foundation of America Publication, No. 9)* (pp. 22): Stuttering Foundation of America.

Johnson, W., Brown, S., Curtis, J., Edney, C., & Keaster, J. (1948). *Speech Handicapped School Children*

New York: Harper & Row.

Johnson, W., & Knott, J. (1937). Studies in the Psychology of Stuttering: I: The Distribution of Moments of Stuttering in Successive Readings of The Same Material. *Journal of Speech Disorders, 2*(1), 17.

Johnson, W., Larson, R., & Knott, J. (1937). Studies in the Psychology of Stuttering: III: Certain Objective Cues Related to the Precipitation of the Moment of Stuttering. *Journal of Speech and Hearing Disorders, 2*(1), 23.

Johnson, W., & Millsapps, L. S. (1937). Studies in the Psychology of Stuttering: VI: The Role of Cues Representative of Past Stuttering in the Distribution of Stuttering Moments during Oral Reading. *J Speech Disord, 2*(2), 101-104.

Kaasin, K., & Bjerkan, B. (1982). Critical words and the locus of stuttering in speech. *Journal of Fluency Disorders, 7*(4), 433-446.

Kalinowski, J., Armson, J., Stuart, A., & Gracco, V. L. (1993). Effects of alterations in auditory feedback and speech rate on stuttering frequency. *Language and Speech, 36*(1), 1.

Kalinowski, J., & Saltuklaroglu, T. (2003). Speaking with a mirror: engagement of mirror neurons via choral speech and its derivatives induces stuttering inhibition. [doi: 10.1016/S0306-9877(03)00004-5]. *Medical Hypotheses, 60*(4), 538-543.

Kelly, E. M., & Conture, E. G. (1992). Speaking Rates, Response Time Latencies, and Interrupting Behaviors of Young Stutterers, Nonstutterers, and Their Mothers. *J Speech Hear Res, 35*(6), 1256-1267.

Kleinow, J., & Smith, A. (2000). Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter. *Journal of Speech, Language, and Hearing Research, 43*(2), 548.

Kloth, S. A. M., Janssen, P., Kraaimaat, F. W., & Brutten, G. J. (1995). Communicative behavior of mothers of stuttering and nonstuttering high-risk children prior to the onset of stuttering. [doi: 10.1016/0094-730X(95)00027-5]. *Journal of Fluency Disorders, 20*(4), 365-377.

Kloth, S. A. M., Janssen, P., Kraaimaat, F. W., & Brutten, G. J. (1998). Child and mother variables in the development of stuttering among high-risk children: A longitudinal study. [doi: 10.1016/S0094-730X(98)00009-6]. *Journal of Fluency Disorders, 23*(4), 217-230.

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. (1989). *Speaking: From intention to articulation*. Cambridge MA The MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(01), 1-38.

Lincoln, M., Packman, A., & Onslow, M. (2006). Altered auditory feedback and the treatment of stuttering: A review. [doi: 10.1016/j.jfludis.2006.04.001]. *Journal of Fluency Disorders, 31*(2), 71-89.

Loucks, T. M. J., De Nil, L. F., & Sasisekaran, J. (2007). Jaw-phonatory coordination in chronic developmental stuttering. *Journal of Communication Disorders, 40*(3), 257-272.

Moore, S., & Perkins, W. (1990). Validity and reliability of judgments of authentic and simulated stuttering. *Journal of Speech and Hearing Disorders, 55*(3), 383.

Newman, R., & Ratner, B. (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.

Oppenheim, G., & Dell, G. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & cognition, 38*(8), 1147-1160.

Packman, A., & Attanasio, J. S. (2004). *Theoretical Issues In Stuttering*: Taylor & Francis.

Packman, A., Menzies, R. G., & Onslow, M. (2000). Anxiety and the Anticipatory Struggle Hypothesis. *Am J Speech Lang Pathol, 9*(1), 88-89.

Perkins, W. (1983). The Problem of Definition: Commentary on "Stuttering". *J Speech Hear Disord, 48*(3), 246-a-249.

Perkins, W. (1990). What is Stuttering? *J Speech Hear Disord, 55*(3), 370-382.

Perkins, W. (1997). Stuttering: why science hasn't solved it. *Nature and treatment of stuttering: new directions, second edition. Boston, MA: Allyn and Bacon*.

Perkins, W., Kent, R., & Curlee, R. (1991). A theory of neuropsycholinguistic function in stuttering. *Journal of Speech and Hearing Research, 34*(4), 734.

Pickering, M., & Garrod, S. (in press). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*.

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

Quarrington, B. (1965). Stuttering as a function of the information value and sentence position of words. *Journal of Abnormal Psychology, 70*(3), 221-224.

Ramig, L. O., Pawlas, A. A., & Countryman, S. (1995). *The Lee Silverman Voice Treatment: A practical guide to treating the voice and speech disorders in Parkinson disease.* Iowa City:: National Center for Voice and Speech.

Rappaport, B., & Bloodstein, O. (1971). The Role of Random Blackout Cues in the Distribution of Moments of Stuttering. *J Speech Hear Res, 14*(4), 874-879.

Schlesinger, I., Forte, M., Fried, B., & Melkman, R. (1965). Stuttering, information load, and response strength. *Journal of Speech and Hearing Disorders, 30*(1), 32.

Schwartz, H. D., Zebrowski, P. M., & Conture, E. G. (1990). Behaviors at the onset of stuttering. *Journal of Fluency Disorders, 15*(2), 77-86.

Sheehan, J. (1953). Theory and treatment of stuttering as an approach-avoidance conflict. *The Journal of Psychology, 36*(1), 27-49.

Soderberg, G. A. (1966). The Relations of Stuttering to Word Length and Word Frequency. *J Speech Hear Res, 9*(4), 584-589.

Stefankiewicz, S. P., & Bloodstein, O. (1974). The Effect of a Four-Week Interval on the Consistency of Stuttering. *J Speech Hear Res, 17*(1), 141-145.

Stemberger, J. P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143-186). Hove, UK:: Lawrence Erlbaum Associates Ltd.

Travis, L. E., Johnson, W., & Shover, J. (1937). The Relation of Bilingualism to Stuttering: A Survey of the East Chicago, Indiana, Schools. *J Speech Disord, 2*(3), 185-189.

Van Riper, C. (1971). *The nature of stuttering*. Englewood Cliffs N.J.: Prentice Hall.

Van Riper, C. (1973). *The treatment of stuttering*: Prentice-Hall Englewood Cliffs, NJ.

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.

Vygotsky, L. S. (1986). Thought and language (rev. ed.). *Cambridge, MA*.

Watkins, K. E., Smith, S. M., Davis, S., & Howell, P. (2008). Structural and functional abnormalities of the motor system in developmental stuttering. *Brain, 131*(1), 50-59.

Weiss, A. L., & Zebrowski, P. M. (1991). Patterns of assertiveness and responsiveness in parental interactions with stuttering and fluent children. [doi: 10.1016/0094-730X(91)90017-7]. *Journal of Fluency Disorders, 16*(2–3), 125-141.

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time-course of phonological encoding. *Journal of Memory and Language, 34*(3), 311-334.

Wingate, M. E. (1967). Stuttering and Word Length. *J Speech Hear Res, 10*(1), 146-152.

Wingate, M. E. (1970). Effect on stuttering of changes in audition. *Journal of Speech and Hearing Research, 13*(4), 861-873.

Wingate, M. E. (1986a). Adaptation, consistency and beyond: I. Limitations and contradictions. [doi: 10.1016/0094-730X(86)90002-1]. *Journal of Fluency Disorders, 11*(1), 1-36.

Wingate, M. E. (1986b). Adaptation, consistency and beyond: II. An integral account. [doi: 10.1016/0094-730X(86)90003-3]. *Journal of Fluency Disorders, 11*(1), 37-53.

Wingate, M. E. (1988). *The structure of stuttering: A psycholinguistic analysis*. Berlin: Springer Verlag.

Wischner, G. J. (1950). Stuttering behavior and learning: A preliminary theoretical formulation. *Journal of Speech and Hearing Disorders, 15*(4), 324.

Wu, J. C., Maguire, G., Riley, G., Lee, A., Keator, D., Tang, C., et al. (1997). Increased dopamine activity associated with stuttering. *Neuroreport, 8*(3), 767-770.

Yairi, E. (1983). The onset of stuttering in two-and three-year-old children: A preliminary report. *Journal of Speech and Hearing Disorders, 48*, 171-177.

Yairi, E., & Ambrose, N. (2005). *Early Childhood Stuttering*. Austin, TX: PRO-ED, Inc. .

Yairi, E., & Ambrose, N. (2012). Epidemiology of Stuttering: 21st Century Advances. *Journal of Fluency Disorders*.

Yairi, E., & Lewis, B. (1984). Disfluencies at the onset of stuttering. *Journal of Speech and Hearing Research, 27*(1), 154.

Figures

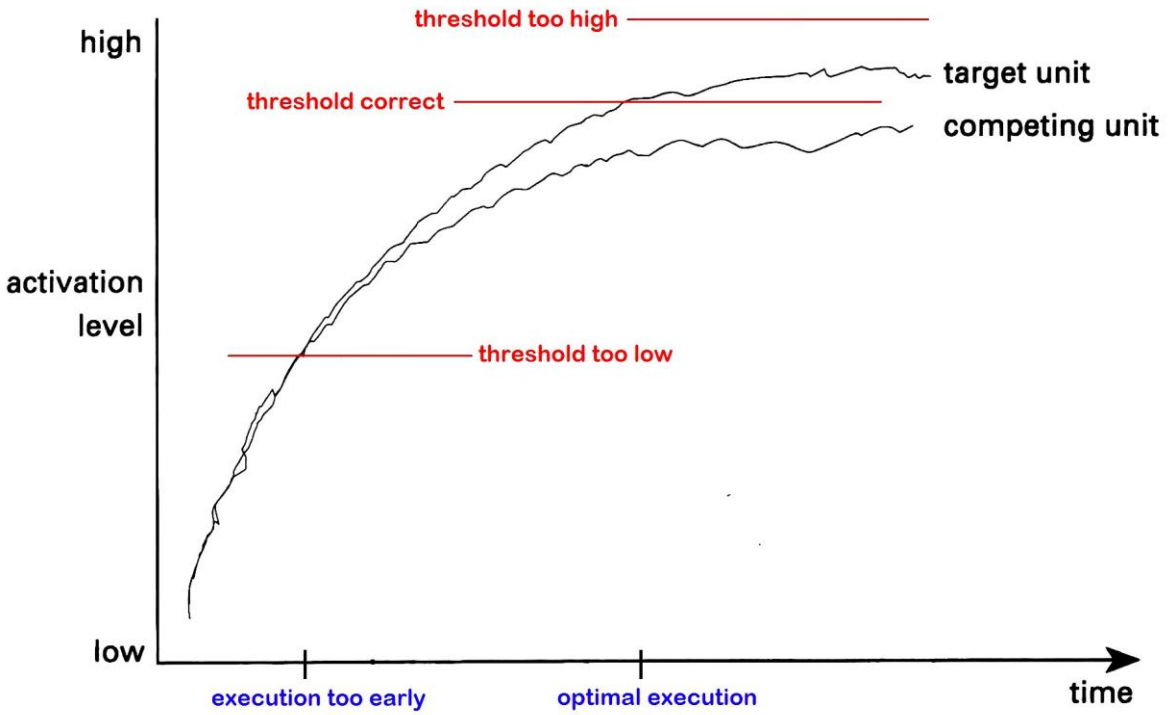

Figure 1. The differing impact on execution of three different release thresholds. If the threshold is too low execution may occur too early, and speakers are likely to select a competing unit in error. If the threshold is too high, execution may not occur at all. When set at an optimal level, selection of the correct unit is more likely than selection of competing units.

Note. Adapted from "Stuttering as a Covert Repair Phenomenon" (p. 188), by A. Kolk and H. Postma, 1997. In R. F. Curlee & G. Siegel (Eds.), Nature and Treatment of Stuttering: New Directions. Copyright 1997 by Allyn & Bacon. Adapted with permission

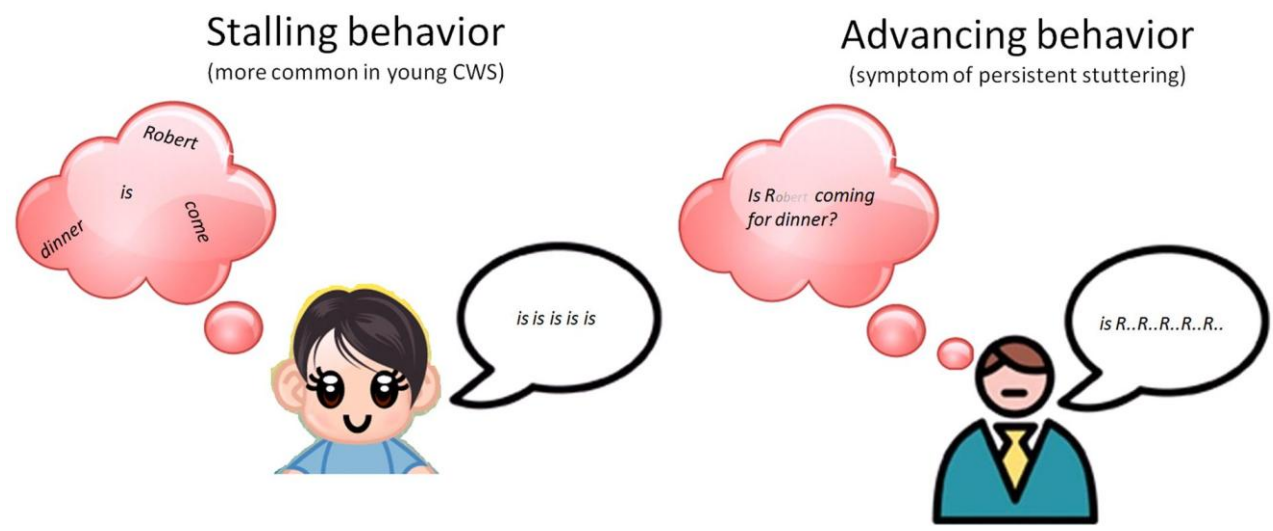

Figure 2. The VRT account of the different causes of stalling and advancing behaviors. In young children who stutter, stallings reflect incomplete syntactic encoding (or lexical access). In the stalling example above, the child does not yet know which word should follow /is/. In contrast, advancings occur after lexical access and syntactic encoding have been completed, but activation of a particular target word is still below the threshold necessary for its execution to be successfully completed. In the advancing example above, the speaker is able to execute the /r/ of Robert, but not yet able to execute the whole word.