



The Tudor Study: Data and Ethics

Nicoline Grinager Ambrose

Ehud Yairi

University of Illinois at Urbana-Champaign

Recent exposure of an experiment (the Tudor Study) conducted in 1939 at the University of Iowa with the aim of studying the effect of verbal labeling on the frequency of disfluency in children who stutter and in normally fluent children has raised strong reactions both from the general public and the scientific community. Allegedly, the investigator and her mentor, a past leader in the field of speech pathology, were successful in their attempts to induce stuttering in normally speaking children; hence, serious accusations of breach of ethics in science have been made. The potential clinical implications of

such conclusions for the treatment of early childhood stuttering are far reaching and negate recent developments that employ direct therapies with preschool children who show signs of stuttering. The purpose of this article is to re-examine the data reported in the Tudor Study and its ethical ramifications. We conclude that none of the experimental questions posed by Tudor and Johnson received empirical support. A broad range of relevant ethical issues is discussed.

Key Words: childhood stuttering, treatment, ethics, The Tudor Study

The general public and the speech-language pathology profession were shocked by a recent report published by James Dryer in the *San Jose Mercury News* edition dated June 11, 2001, regarding the experimental work directed by the late Professor Wendell Johnson of the University of Iowa. Dryer alleges that Johnson, one of the most legendary figures in the history of the field of communication disorders, who had enormous influence on theory, research, and treatment of stuttering until his death in 1965, engaged in scientific misconduct while directing the master's thesis of his graduate student, Mary Tudor, a thesis later referred to as the "Monster Study." The reporter contends that the thesis project was essentially an experiment designed to induce stuttering in normally fluent children. He suggests that the underlying motivation was to test Johnson's diagnosogenic theory that negative reactions to normal speech disfluencies cause stuttering in children. It only added to the report's emotional impact that the experiment, carried out in 1939 over a period of 3 to 4 months, involved orphans at the Soldiers and Sailors Orphans' Home in Davenport, Iowa. Dryer also notes that Johnson apparently suppressed the existence of the study in light of the World War II abuses of human subjects by Nazi scientists and physicians.

According to Dryer, the Tudor study successfully achieved its objectives. Specifically, he concludes, "In the experiment, Tudor had subjected half of the children to criticism to make them self-conscious about their speech,

eventually driving most of them to stutter." Six of 22 children who were normally fluent, received negative evaluative labeling regarding their disfluencies. Another 5 children who were allegedly already stuttering also received such labeling. Referring to a now 77-year-old woman, the author states that, "At age 15, ...a normal speaker, was induced to stutter." Referring to yet another participant, the reporter states that "At age 6, she had spoken fluently, but was induced to stutter in the experiment." The report thus suggests that the children developed chronic stuttering that was impossible to reverse, indicating that the children were not only induced to stutter but became people who stuttered. In fact, of the 22 children in the study, 12 were judged to be normally fluent, and 6 of these received negative evaluative labeling regarding their disfluencies. Ten were judged to already be stuttering before the study, and 5 of these children received negative labeling. Reportedly, Ms. Tudor (now Mrs. Tudor Jacobs), age 84, is convinced that her experiment did indeed induce persistent stuttering. Three of the study's participants have also recently come to hold Tudor Jacobs and the late Professor Johnson responsible for their life-time speech handicap.

The *Mercury News* report brings extremely important ethical and scientific issues to the forefront. In an interview broadcast on many National Public Radio stations, Dryer stated that, despite its ethical lapses, the study was characterized by good design and that the results supported the diagnosogenic theory. These statements, as will be

demonstrated, are not backed by scientific reviews of the study. Additional reactions that appeared soon after the *Mercury's* report, mainly critical of Johnson's theory and/or ethics, were expressed on internet websites and in newsletters of stuttering support and professional groups (e.g., Ratner, 2001). The present authors also published a brief letter in *The ASHA Leader*, the news magazine of the American Speech-Language-Hearing Association (Yairi & Ambrose, 2001) that challenged Dryer's conclusions. However, debate continues. More recently, Erard (2001) in an article published in *Lingua Franca*, stated that "Johnson, rather sinisterly supposed that the best way to win support for his theory was to turn experimental subjects into stutterers" (p. 40). In view of the serious nature of the issues raised by Dryer, we feel that these issues deserve in-depth scrutiny by the scientific community. It is time to take a close look at the Tudor study and its data.

The Theory

The initial exposition of the diagnosogenic notion appeared briefly in a scientific article concerning the onset of stuttering published by Wendell Johnson in 1942. The "theory," borrowing heavily from ideas generated by scholars of general semantics (e.g., Korzybski, 1941) was intuitively appealing and quickly became accepted by many speech pathologists around the world. Its general concepts also were adopted by other child health and educational professionals. These concepts also became familiar to the general public, especially to parents who became concerned about their young children's stuttering. Its primary tenet was that the mere diagnosis of a young child's speech as "stuttering" can cause the disorder. As the theory goes, a child exhibiting normal speech disfluencies typical of preschoolers is given a diagnosis of "stuttering" by overly anxious parents, who then react to the normal disfluencies as though they were abnormal or stuttered. By and large, however, such diagnoses are erroneous, reflecting parents' lack of knowledge regarding norms of children's speech. As parents react in various negative ways, including disapproving comments, instructing the child to slow down or start again, or even showing concerned facial expressions, children attempt to avoid the unwanted normal disfluency by changing their speech patterns. Trying hard not to repeat, children become anxious and tense so that simple, easy, "normal" repetitions become tense and gradually turn into the struggle behavior that is stuttering.

The Tudor Study

Tudor's (1939) thesis, titled "An Experimental Study of the Effect of Evaluative Labeling on Speech Fluency," was designed to test several assumptions of the diagnosogenic theory. Specifically, Tudor and Johnson investigated the influence of applying the labels "stuttering" and "normal speaker" to children on the frequency of disfluencies in normal speakers and those who already stutter.

Objectives

The study was designed to answer the following questions (Tudor, 1939):

1. Will removing the label "stutterer" from those who have been so labeled have any effect on their speech fluency?
2. Will endorsement of the label "stutterer" previously applied to an individual have any effect on his speech fluency?
3. Will endorsement of the label "normal speaker" previously applied to an individual have any effect on his speech fluency?
4. Will labeling a person, previously regarded as a normal speaker, a "stutterer" have any effect on his speech fluency? (p. 2)

As can be seen, the objectives of the study emphasize the influence of labels on the degree of speech fluency. They do not explicitly indicate attempts to induce stuttering.

Participants

To test the theory, Tudor had 256 residents screened at the Soldiers and Sailors Orphans' Home in Davenport, Iowa. Screenings were conducted by five judges trained in speech pathology, who listened to each child's speech for three minutes and rated it on a 5-point scale, with 1 being the lowest degree of fluency, 3 representing average fluency, and 5 the highest degree of fluency. Judges were instructed to rate about 40% of the children as 3, 20% as 2, 20% as 4, 10% as 5 and 10% as 1. To this group were added children judged to stutter by the institution's staff. From this pool, 22 children were selected for the study. Ten were judged to stutter, and twelve were normal speakers chosen at random from the remaining children. Participants ranged in age from 5 to 16 years ($M = 11$ years). Handedness and eyedness measures were administered by the investigator and IQ was determined, but neither the test nor source of information were specified. The mean IQ for all 22 participants was 85 (range: 61 to 122). Eleven, or half of the children, scored below average, 10 fell within normal limits, and 1 was in the above average range.

Prior to the experiment, each subject read aloud for 5 minutes and spoke spontaneously for 5 minutes. Each of the five judges assigned a fluency rating and gave a description of the child's speech. The samples were audio recorded for later transcription. The experimenter, Ms. Tudor, then interviewed each child and assigned her/him to one of the four experimental groups. The 10 children judged to stutter were divided into Groups IA and IB, and the 12 children judged as normally fluent were divided into Groups IIA and IIB.

Experimental Groups

Group IA included 5 children who stuttered: all were males, with a mean age of 12.4 years (range: 10–14) and a mean IQ of 86.8 (range: 70–122). During treatment, they were told that they were not stutterers.

Group IB included 5 children who stuttered: 2 males, 1

female, and 2 gender unknown, with a mean age of 12 years (range: 6–15) and mean IQ of 86.8 (range: 81–91). They continued to be labeled as stuttering.

Group IIA included 6 normal speakers: 4 females and 2 males, with a mean age of 10.67 years (range: 5–15) and a mean IQ of 84.5 (range: 65–99). They were told that they stutter and should do anything possible to avoid stuttering.

Group IIB included 6 normal speakers: two males, 1 female, 2 gender unknown, with a mean age of 10.67 years (range: 6–16) and a mean IQ of 85.45 (range: 61–99). They were told they had good speech.

Treatment

Each child was seen by the experimenter for several sessions between January or February and May, 1939. The role of the judges in these sessions is unclear. Tudor states that the judges endorsed or removed labels, which they did not do in their initial severity rating, indicating that judges must have conducted additional observations during the treatment phase. Children in Group IA were seen 3 or 4 times. For Groups IB and IIB, no mention of the number of session is made and only initial and final sessions are discussed. Children in Group IIA (normal speakers who were told that they stutter and should avoid it) were seen 8 to 9 times. During the initial interview, the experimenter read instructions to each child. Instructions were different for the children in each group according to the objectives of the study. Children in Group IA were told that many children also had similar trouble in speaking and that they should pay no attention to it, as they would soon “be able to speak well” (p.9). Children in Group IB were asked a number of question about various features of their stuttering. Following are the specific instructions given to Group IIA, the normal speakers to whom the label “stutterer” was applied (Tudor, 1939):

The staff has come to the conclusion that you have a great deal of trouble with your speech. The type of interruptions which you have are very undesirable. These interruptions indicate stuttering. You have many of the symptoms of a child who is beginning to stutter. You must try to stop yourself immediately. Use your will power. Make up your mind that you are going to speak without a single interruption. It's absolutely necessary that you do this. Do anything to keep from stuttering. Try very hard to speak fluently and evenly. If you have an interruption, stop and begin over. Take a deep breath whenever you feel you are going to stutter. Don't ever speak unless you can do it right. You see how (the name of a child in the institution who stuttered rather severely) stutters, don't you? Well, he undoubtedly started this very same way you are starting. Watch your speech every minute and try to do something to improve it. Whatever you do, speak fluently and avoid any interruptions whatsoever in your speech. (pp. 10–11)

Children in Group IIB were told that they spoke very well.

According to the experimenter, the scripts served only

as general guidelines with the language modified to make it comprehensible to each child with whom an interview was held. During the session, the experimenter implemented the instructions while conversing with the child. For example, when a child in Group IA (whose stuttering was labeled as normal disfluency) talked about his stuttering, the investigator responded saying that he did not stutter and probably repeated words because he was not sure what he was going to say. On the other hand, when a child in Group IIA (whose normal disfluency was labeled as stuttering) exhibited disfluency, the experimenter responded by saying that it was stuttering and using remarks as stated in the instructions above.

In addition, the orphanage staff were told that Group IA didn't really stutter and that they should not pay any attention to or remark on the children's speech, and that Group IIA stuttered and that they should watch the speech of those children, interrupting them and having them repeat whenever they stuttered. Tudor, however, states in her thesis that the caretakers later admitted they didn't follow these instructions because of their “low esteem” for the children.

Measures

Two measures were used, each taken at pre-and posttreatment (these time periods are referred to as “beginning” and “end” in the thesis):

1. Five judges independently rated the severity of stuttering for each child at the beginning and end of the experiment on a 5-point scale with 1 as the lowest degree of fluency, 3 as average, and 5 as the highest degree of fluency. The judges were unaware of the treatment group to which each child was assigned.

2. The frequencies of speech disfluency types transcribed from the recorded speech samples were calculated, including syllable repetitions, word repetitions, phrase repetitions, interjections, and pauses. No further definitions of the disfluencies are provided. Thus, no distinctions between disfluency types typical of stuttered speech and those typical of normally fluent speech are made. Conventionally, of the five types identified, syllable repetitions can be considered the hallmark of stuttering. Monosyllabic word repetitions are also prominent in stuttered speech, but the category as used in the study included words regardless of number of syllables. Prolongations or blocks are not mentioned or tabulated in the disfluency analyses. Phrase repetitions are typically regarded as normal disfluencies, and interjections are the most common disfluency type in the speech of normally speaking people, although both may also be used to avoid or postpone stuttering. Pauses can also indicate hesitancy in avoiding or postponing stuttering, but they are also very common in normally fluent speech. Tudor refers to pauses between sentences (p. 69), and inspection of the transcripts shows that pauses were marked between words, phrases, or sentences, making it clear that pauses do not represent blocks. Of all disfluency types then, only syllable repetitions and a portion of the word repetitions can be considered most likely to reflect behavior typically

perceived as stuttering. In addition to quantified data, the investigator took notes from the experimental sessions. There is no information about the length of the sessions, although they appear to have varied in length, judging from the notes.

Reanalyses of the Data Reported in the Tudor Study

To perform our analyses, we entered participant data into a file, including group, gender, IQ, disfluency frequencies for all five types for initial and final speech samples, and initial and final judges' severity ratings. These were taken directly from the text and tables in the Tudor thesis and were not modified in any way. The document itself presents no statistics other than group means, and no statistical tests to evaluate differences in group means.

Perceptual Data

Fluency severity ratings obtained from judges prior to and following treatment were used to calculate means for each of the experimental treatment groups (see Table 1). Recall that a score of 1 = very disfluent, 3 = average, and 5 = very fluent. Three striking observations emerge from these data. First, at the beginning of the experiment, there appear to be almost no differences in the mean ratings of children in the four groups. Thus, no reasonable differentiation between children who stutter (Groups IA and IB) and normally speaking children (Groups IIA and IIB) may be made. Second, for each group, there is no meaningful difference between the pre- and posttreatment severity ratings. Third, there are no meaningful differences among the groups at the end of the study. A repeated measures analysis of variance revealed no significant differences between the beginning and end ratings [$F(1, 17) = .187, p = .671$], and no significant interaction between the ratings and the four groups [$F(3, 17) = .195, p = .898$]. The trends for each group are not only insignificant, but are also in the wrong (unexpected) direction for most. Group IA (stuttering labeled as normal) had decreased

fluency, and Group IIA (normally fluent children targeted for stuttering) had improved fluency. Both Groups IB, (stuttering labeled as stuttering) and IIB (normally fluent labeled good speakers) improved slightly in fluency.

In Group IA (stuttering label removed) the range of severity ratings pretreatment is wide. Two participants fall at the more disfluent end of the scale, at 1.4 and 2.3, one falls in the middle at 3.2, but two are rated at the fluent end of the scale at 4.0 and 4.2. For these last two, and perhaps the child rated at the middle of the scale, it appears that the judges did not believe that these children stuttered, although the institution staff felt that they did. The two lowest rated children did increase their fluency by the end of the study. The three other children, however, who began well in the fluent range, decreased their level of fluency by the end of the study. In Group IB, four of the five children were initially rated in the 2.0–2.8 range, and of these, at the end of the study, two had increased fluency and two had decreased fluency. The remaining child was rated on the fluent side of the scale at 3.6 at the beginning and was rated about the same at the end. Group IIB (normally fluent speakers) were rated from 2.3 to 3.4, all falling within the range of ratings for all of the 10 stuttering children in Groups IA and IB. Their ratings remained fairly stable, with only one showing a distinct increase in fluency.

Individual fluency ratings for Group IIA (the six normally speaking children in whom the investigator attempted to increase disfluency) and summaries of the statements of the five judges, are provided in Table 2. None of the participants received initial fluency ratings of 5, or even four, which would place them at the high ("very fluent") end of the scale. The highest fluency rating for any participant in Group IIA was 3.1. At the termination of the study, one participant's (#11) rating dropped by 0.6 scale points (more disfluent) while another participant's (#13) rating improved by 0.6 points (more fluent). The greatest improvement in fluency was shown by Participant #15, whose perceived fluency improved by 0.8 scale points. Notably, this participant was among those reported in the *Mercury News* to have developed stuttering as a result of the study. Additionally, it should be noted that, posttreatment, none of the children received even a single description by the independent judges indicating stuttering. It is partially on the basis of the perceptual data addressed here that Dryer concluded—and reported that Tudor concluded—that stuttering was established as a result of the study's methods. Their conclusion is clearly not supported by the data.

Disfluency Data

Next, we analyzed the disfluency data from the recorded speech samples from all of the children in each of the four experimental groups. The pre- and posttreatment group means for each disfluency type are presented in Table 3. A repeated measures MANOVA indicated no significant differences pre- and posttreatment [within-participants factor, $F(5, 14) = 2.61, p = .07$], for any of the groups [between-participants factor, $F(15, 48) = 0.88, p = .59$], for syllable repetitions, word repetitions, phrase repetitions,

TABLE 1. Mean (and standard deviation) for judges' severity ratings before and at the end of the study for the four groups.

	Ratings	
	at Beginning	at End
Group IA Stuttering labeled normal	3.02 (1.18)	2.92 (0.82)
Group IB Stuttering labeled stuttering	2.62 (0.64)	2.80 (0.75)
Group IIA Normal disfluency labeled stuttering	2.83 (0.24)	2.92 (0.68)
Group IIB Normal disfluency labeled good speech	2.88 (0.44)	2.92 (0.23)

Note. Ratings; 1 = least fluent; 5 = most fluent.

TABLE 2. Ratings and judges' reports for children in Group IIA.

Subject Number	Rating at Beginning	Reports of 5 Judges at Beginning of the Study	Rating at End	Reports of 5 Judges at End of the Study
11	2.6	2 reported normal speech 3 noticed repetitions	2.0	4 reported no stuttering 1 said "difficult to get to talk"
12	2.8	2 reported adequate speech 2 reported hesitations 1 reported a few repetitions	2.8	3 reported no stuttering No report for 2
13	3.0	3 reported normal speech 1 jerky speech 1 occasional repetitions	3.6	4 reported no stuttering 1 reported "hesitant" speech
14	2.5	1 reported adequate speech 4 reported repetitions Hesitancy and interjections also mentioned	2.5	4 reported no stuttering 1 reported "hesitant" speech
15	3.0	5 reported repetitions	3.8	4 reported no stuttering 1 reported speech was shaky but did not bother subject.
16	3.1	2 reported normal speech 3 reported pauses or interjections	2.8	3 reported no stuttering; 1 said 1 said "answered briefly" 1 reported speech "hesitant, seems embarrassed"

interjections, and pauses. Nor was there significant interaction of time by group [$F(15, 48) = 1.20, p = .31$]. Although power may be inadequate to detect differences for this sample, the lack of any apparent trends makes these results appear sound. Inspection of Table 3 reveals that the greatest change occurred for interjections, which increased in Group IIA, and for pauses, which increased in Group IIA and decreased in Group IIB. As noted previously, interjections are typically regarded as a normal speech phenomenon, although they also may be used as a postponement or avoidance technique by some individuals who stutter. Pauses may indicate many things, such as thinking, distraction, hesitancy, or postponement or avoidance. It is not possible to say if these changes bear any relation to stuttering in this study. As described by Tudor, they do not appear to include tense pauses. Syllable repetitions, a definite sign of stuttering, indicate no considerable change. Again, Dryer's conclusion that stuttering was induced as a result of the study is not supported by Tudor's own data.

Figure 1 illustrates the individual disfluency data for members of Group IIA (normal speakers in whom an attempt was made to induce stuttering). The solid lines in each individual figure represent the trends of the two disfluency types most likely to be judged as stuttering (i.e., part word and whole word repetitions), while the dashed lines indicate normal types of disfluencies (i.e., pauses, phrase repetitions, and interjections). Part-word repetition is initially at a level generally considered as stuttering (5.23), and becomes greatly reduced for participant 11. It is slightly reduced for children 12 and 14, and increases slightly for children 13, 15, and 16. Word repetition increased slightly for two children and remained about the same for the other four. All six children show an increase in interjections, and children 11, 12, 14, and 16 show a distinct increase in pauses.

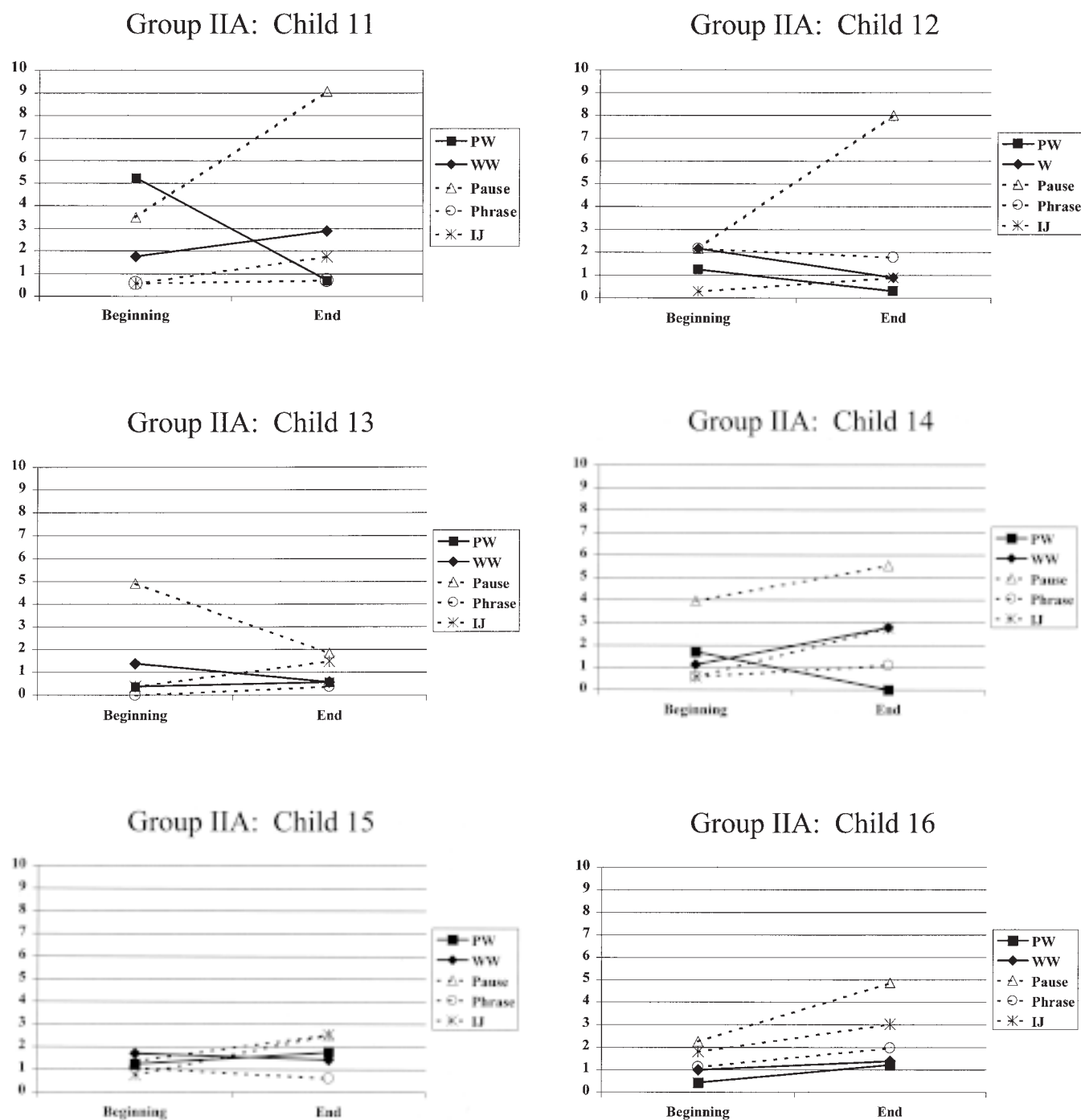
Clinical Notes

As mentioned in the Method section, the investigator took notes during the experimental sessions. The initial interview following judges' ratings was also the first treatment session, and the notes reflect children's reactions and responses. A synopsis is presented here.

TABLE 3. Group means for disfluency data.

Group/Measure	Means-Before	Means-End
IA – Stutterers labeled normal speakers		
Syllable repetition	1.18	0.84
Word repetition	3.74	2.09
Phrase repetition	1.76	0.76
Interjections	2.94	3.21
Pauses	5.93	3.58
IB - Stutterers labeled stutterers		
Syllable repetition	3.56	2.03
Word repetition	1.75	2.16
Phrase repetition	0.42	0.93
Interjections	2.39	3.50
Pauses	5.33	4.77
IIA - Normal speakers labeled stutterers		
Syllable repetition	1.70	0.76
Word repetition	1.52	1.65
Phrase repetition	0.93	1.09
Interjections	0.74	2.08
Pauses	3.03	5.32
IIB – Normal speakers labeled good speakers		
Syllable repetition	0.49	0.41
Word repetition	1.75	1.40
Phrase repetition	1.00	0.51
Interjections	0.60	0.49
Pauses	4.67	2.75

FIGURE 1. Changes in disfluency types from beginning to end for individual children in Group IIA.



Participant #11, a supposedly normally speaking child who was targeted for increase in disfluency (Group IIA), had part-word repetitions of 5.23%, a percentage consistent with that typically exhibited by children who stutter, at the beginning of the program (Ambrose & Yairi, 1999). It is not clear why this child was assigned to the normally speaking group. During the first interview, when asked if

she knew anyone who stuttered, "she said that she did." It is not clear if she meant "yes, I stutter" or "yes, I know someone who stutters" (p .65). The notes indicate that during one session in the middle of the program she exhibited some prolongations. None are mentioned toward the end of the program. Notably, although prolongations are mentioned as a type of disfluency in the notes for this

participant, it is not one of the disfluency types included in the disfluency analyses. The investigator's summary for the participant states: "Considering all of the data obtained at the beginning, during, and at the end of the experimental period, this participant showed a decrease in fluency as rated and an increase in the percent of speech interruptions. During the experimental period her manner of speaking changed markedly. At the beginning of the period she spoke freely and connectedly, but at the end of the period she was very unwilling to talk and her responses consisted of single words or short phrases" (p. 116). The reported increase in pauses support this claim, but the reduction of syllable repetitions do not.

Participant #12 also "said that she did" (p. 74) when asked if she knew anyone who stuttered, and again, it is not clear what is meant. The notes from the interviews indicate that the number of speech interruptions went up and down throughout the study. Her speech was described as choppy and slow, and after she was told she was stuttering during the first session, she began to frequently place her hands over her mouth to conceal her interruptions. There were no descriptions of stuttering-like behavior in the notes. At the end of the program her teacher described her as "restless in the class and that whenever she recited her face became flushed" (p. 80). The investigator's summary states: "This subject's fluency rating remained the same, but she showed an increase in the percent of speech interruptions. At the beginning of the period she had been willing to speak in any situation, but at the end of the period it was practically impossible to get her to speak in an unfamiliar situation" (p. 116).

For Participant #13, notes from the first session state: "I asked him if anyone had ever told him that he had trouble with his speech. He replied that no one ever had. ...Did you ever catch yourself stuttering? Yes, I think so...once in a while. You have a difficult time speaking don't you? Nodded head" (p. 83). The use of such leading questions makes interpretation of the child's responses difficult. The notes also indicate that in the first session he prolonged the "f" in the word "fractions." The following notes taken during the middle of the program stated that: "...he had no significant speech interruptions. The only noticeable phenomenon of stuttering which occurred was a breath-iness in reading" (p. 85). At another session during the program, he was noted to stutter and exhibit prolongations. The investigator's summary states: "This subject showed an increase in speech fluency as rated and a decrease in the percent of speech interruptions. He is the only subject in this group who decreased the percent of speech interruptions. He is however, very conscious of his speech and at the end of the period he spoke more slowly and carefully than he did at the beginning of the period" (pp. 116-117).

Participant #14. The notes do not provide any indications of stuttering during the entire the program. The investigator's summary states: "This subject's fluency rating remained the same, but there was an increase in the percent of speech interruptions. At the end of the experimental period it was almost impossible to get the boy to speak in a situation other than play, whereas at the beginning of the period he spoke

freely in a new situation. In school his grades were markedly lower during the last six-week period" (p. 117).

For Participant #15, the notes from the first experimental session state that "she was very easily influenced. She seemed to enjoy the fact that there was something different about her for she craved attention and her speech was noticed by other people" (p. 101). Her speech was described as containing repetitions and interjections. Tudor states that at the next session, she had forgotten all about her "stuttering." Toward the middle of the program some prolongations were noted but no stuttering is mentioned, although, in responses to the experimenter's question about whether she stutters, the participant answered positively. During the last three sessions no prolongations or stuttering were described. Furthermore, the participant said that her best friend did not know that she stuttered. According to the experimenter, "Unlike the other children in this group she was never unwilling to speak, although her speech interruptions increased consistently throughout the experimental period. ...seems to enjoy the attention that I gave her concerning her speech" (p. 105). The investigator's summary paints a somewhat different picture than the notes, stating: "This subject increased in speech fluency as rated, but she also increased the percent of speech interruptions. This girl's speech had, in increasing number, repetitions, prolongations, and interjections throughout the period. She also ceased to tell stories or converse freely with her playmates" (p. 117).

For Participant #16, the clinical notes from the first session state that "...she repeated and I called her attention to it. And told her it was stuttering. I told her that stutterers also prolonged sounds and hesitated before they say words. She immediately picked up all of these interruptions in her own speech." In the third session, approximately 6 weeks later, the following is reported: "How is your stuttering? Seems like I get worse all the time. Whenever I go to say something, I just can't say them." It seems that the participant adopted the experimenter's language referring to all disfluencies as stuttering, but this does not mean that her disfluency behaviors changed and became stuttering. During the remainder of the session, the girl continued to report about her stuttering. The clinician's description of her speech, however, points out that the girl would laugh or cover her mouth when she had interruptions but it does not mention stuttered speech. At only a single instance, when she read a story about "Dodo," the girl kept repeating "Do-Do-Do-Do" (p. 114), but it is not clear if this was playful behavior, syllable repetition typical of stuttering, or a simple repetition of the name "Dodo." The investigator's summary states: "This subject showed a decrease in fluency as rated and an increase in the percent of speech interruptions. During the experimental period she developed mannerisms characteristic of some stutterers, such as 'snapping' her fingers to 'get a word out'" (p. 117).

Summary of Children in Group IIA

The severity rating for Participant #11 went from 2.6 to 2.0, a decrease in fluency. Pauses increased dramatically, and part-word repetitions decreased markedly. Initially, 3

judges reported repetitions, and 2 normal speech; none of the judges reported stuttering at the end of the study. Notes indicate that she stuttered for at least some of the sessions, and that her fluency decreased and her interjections increased. She used to speak freely, but at the end of the study she was unwilling to talk. Her assignment to the initially normally fluent group is questionable. We also question existence of stuttering at the end, although it seems that she was reacting negatively in her communication attitude and behavior. Given her age and gender, she can be identified as one of those described in Dryer's newspaper article as being induced to stutter. As stated above, however, it is possible that she entered the study already stuttering, given the number of syllable repetitions, and that she continued to stutter for at least a few sessions, as reported by the investigator in her experimental notations. This child was reported by Dryer to have been induced to stutter by the study.

For Participant #12, severity remained the same at 2.8 in both the beginning and end sessions. The data show that her pauses increased markedly. The notes indicate that disfluency frequency varied. Speech was choppy, and she reacted negatively to her own disfluencies, although no stuttering-like speech behaviors are noted. She previously had spoken freely, but at the end she was unwilling to talk in unfamiliar situations.

Participant #13 became more fluent over the study, with a severity rating improving from 3.0 to 3.6. Pauses decreased to a moderate extent. At the first interview, he stated that he occasionally stuttered. The notes also indicate the presence of some features of stuttering at various sessions, and the absence of interruptions at another. A prolongation is reported during the initial interview. His fluency is reported to have increased, and he is the only child in the group whose frequency of disfluencies decreased. In spite of this, he was reported to speak slower and more carefully at the end of the study. This child, then, was described as exhibiting some stuttering-like behaviors at the beginning of the study, yet he was assigned to a "normally speaking" experimental group. Some possible stuttering is also described during the study, although not at the end.

Participant #14's initial and final severity ratings were also identical. Interjections and pauses increased, and syllable repetition decreased. There is no mention of stuttering in the notes except that his interruptions increased, and he became unwilling to speak. Again, it appears that his communication became inhibited, but stuttering did not appear to be exhibited.

Fluency severity ratings for Participant #15 increased from 3.0 to 3.8, the greatest improvement in the group. There was very little change in disfluency frequencies. Notes mention repetitions and interjections, and some prolongations later in the study. The author's summary states that repetitions, prolongations, and interjections increased over the study, and later in the study the child reported that she herself stuttered. In one section she is described as very willing to speak, but in another, as no longer conversing freely. There is an apparent contradiction between the report of increases in interruptions by the author and improved fluency by the judges. According to

the newspaper report, this participant stuttered as a result of the experiment.

Participant #16's fluency rating decreased only slightly from 3.1 to 2.8. Frequency of disfluencies increased slightly for all types. Notes indicate that she considered herself to stutter as the study progressed, and that she reacted strongly to her own disfluencies by covering her face and laughing or by snapping her fingers. Again, the data, notes, and summary provide contradictory information. This child was reported by the newspaper to have been induced to stutter by the study.

In summary, according to the investigator's notes, it appears that two of the children (11 and 13) may have exhibited some stuttering at the beginning of the study; three (11, 12, 14,) indicated communication inhibition, but not stuttering, at the end of the study; and 2 (15 and 16) were described by Tudor as developing features of stuttering. The data, however, do not confirm development of stuttering, but are consistent with the advent of low self-esteem regarding speaking abilities.

Discussion

Experimental Methods

In general, the experimental methodologies employed in the Tudor study reveal weaknesses in design and execution. These will be discussed prior to commenting on the results.

Participants

The sample chosen for the study cannot be considered to be representative of the population of children (stuttering or fluent) at large. First, the children were raised without parents in an institution. Second, the IQ level of the participants was below the population average. This may be a significant confounding factor, in that a number of past investigators have found that groups with lower than average IQ tend to exhibit higher incidence of stuttering than the population at large (for a review, see Bloodstein, 1995) or higher levels of disfluency even if they are not stuttering (Otto & Yairi, 1976).

The fact that two of the six children in Group IIA, the normally speaking children targeted for increased disfluency, initially showed evidence of stuttering is alarming in relation to the process of participant selection. During the initial interview, one child stated that he stuttered. It would seem to be rather improbable that his statement reflected the influence of treatment from the beginning of the first experimental treatment. The amount of speech disfluency documented in the speech of another member of Group IIA, Participant #11, clearly placed her within the conventional classification of stuttering at the beginning of the study. Her speech contained up to almost 7% stuttering-like disfluencies. Of these, 5.2% were syllable repetitions, the most common disfluency type in speech of children who stutter (Ambrose & Yairi, 1999; Yairi & Lewis, 1984). Typically, 3% is regarded as sufficient for suspecting stuttering (Conture, 2001). To conduct a clean study, those children should have been disqualified.

Thus, Dryer's claim that Participant #11 began stuttering as a result of the experimental procedures has no acceptable backing and should be questioned in light of the possible presence of a previously existing stuttering problem. This fact was apparently overlooked both by the investigator and the reporter. As shown in Table 1, the study failed to initially establish acceptable distinction among all four of the groups. For example, two children in Group IA, classified as stutterers, received initial fluency ratings that were appreciably better than those of any of the "normal" participants in Group IIA.

Procedures

Numerous questions may be raised about the experimental procedures employed in the Tudor thesis. Although it appears from the notes that the investigator conducted individual sessions with each of the 22 participants, the thesis does not contain a clear statement in this regard. Could a single clinician have administered all the sessions, keeping them equal in length, within the short time available? From the notes, it is clear that sessions varied in length. Additionally, as indicated above, participants did not receive equal numbers of treatment sessions. According to the statements in the study, children in Group IIA received between 8 and 9 sessions, whereas Group IA received either 3 or 4 sessions. No information is given for Groups IB and IIB. As the investigator was very familiar with the assumptions of the study, bias in applying equal treatment, interpreting participants' responses during the sessions, and in analyzing the recorded speech, may also have been a factor.

As noted earlier, Tudor considered all of the measured disfluencies to be reflective of stuttering, referring to them as "interruptions." Thus, when her notes indicate increase in the percent of interruptions during the experimental treatment, it is very probable, based on the data, that much of it is normal disfluency, as we presently understand it. The treatment of very diverse speech phenomena as if they were all equally relevant to stuttering is unfortunate. Significantly, it was that indefensible approach that allowed Johnson et al. (1959) to forge "overlap" between disfluency of those who stutter and normally fluent children, thus confusing the issue. Interpretation of the results is further hindered by the use of inconsistent terminology for disfluency types. As already mentioned, the investigator refers to "prolongations" in her clinical notes. Prolongations are clear signs of stuttering. However, prolongation is not one of the disfluency types included in her formal analyses of the recorded speech. Were the children, for some reason, producing prolongations only in her experimental/clinical sessions but not in other situations where they were recorded? Another example of loose and potentially misleading terminology is the application of "stuttering" to characteristics such as "breathiness in reading."

Another weakness in the administration of the experimental treatment, as planned, is the failure to carry on with the staff's follow-ups. Because of the small number of sessions, spread over 3 to 4 months, staff participation was deemed important. Thus, the potential for round-the-clock

monitoring appears to be an important motivation for conducting the study in a residential institute. As Tudor reported, however, "They apparently felt that it was a waste of time to give these children special attention. In any event, the suggestions offered were, for the most part, disregarded. As it was, the children received their stimulation entirely from the writer" (p. 147). She then concludes: "If these children had constantly been reminded of their speech they would have undoubtedly reacted more positively" (p. 147). As far as the experimental outcome is concerned, such comments provide strong indications that the investigator had reservations about the results not matching expectations. Further, brief speech and reading samples were obtained from each participant at the pre- and posttreatment recordings, and none were taken during the treatment period itself. Although more recent investigators have also employed speech samples as short as 200 to 300 words in various research studies (Conture & Kelly, 1991; Yaruss, 1997), including samples as short as one minute in clinical treatment studies (Onslow, Costa, & Rue, 1990), there are reasonable grounds to suspect that short samples may not adequately represent the full range of a person's disfluency (Yairi, 1997). And, finally, the clinical notes and summaries of treatment results make frequent reference to the impact of the treatment on the emotional reactions and various related behaviors of the participants. The Dryer report, Ratner (2001) and Erard (2001), all make references to these reactions as consequences of the experimental procedures. There is no support for these claims and no hard data. These conclusions are thus lacking substantiation for the simple reason that no baseline nor posttreatment data were collected for any of these parameters. We do not know how these children behaved and reacted prior to and after the experiment.

Reliability and Statistical Evaluation

No reliability data are provided for any of the measures employed in the study. Additionally, neither the investigator's observations of the children's speech nor their reactions during the sessions were verified by other observers. As already mentioned, no statistical analyses were performed on the data reported in the Tudor thesis to offer confidence in their meaning.

Results

Our reanalysis of the original data from the Tudor study failed to confirm the hypotheses of the investigator and her advisor. Differences in judges' perceptual ratings pre- and posttreatment were statistically nonsignificant. Similarly, mean differences in frequencies of specific disfluency types pre- and posttreatment were nonsignificant. For the disfluencies counted in this study, there were no observable or significant increases or decreases in any disfluency type for any of the four groups. The only larger change was for the normally fluent group labeled stutterers (Group IIA) who increased interjections in their speech, a disfluency type that is typically regarded as normal. Some individuals in this group had considerable change in one or two types,

but none that provided direct evidence of stuttering. This group of six children is of primary interest because they received negative comments about their normal disfluencies, and have complained, according to Dryer's investigative report, about long-term serious problems related to fluency.

As we have already noted, for Group IIA (normal speakers labeled as stuttering), none of the judges described any of the children's speech as stuttered at the end of the study. This conclusion is reflected in Tudor's final statement: "All of the subjects in Group IIA (the children to whom the label "stutterer" had been attached) showed similar types of speech behavior during the experimental period. A decrease in verbal output was characteristic of all six subjects; that is, they were reluctant to speak and spoke only when they were urged to. Second, their rate of speaking was decreased. They spoke more slowly and with greater exactness. They had a tendency to weigh each word before they said it. Third, the length of response was shortened. The two younger subjects responded with only one word whenever possible. Fourth, they all became more self-conscious. They appeared shy and embarrassed in many situations. Fifth, they accepted the fact that there was something definitely wrong with their speech. Sixth, every subject reacted to his speech interruptions in some manner. Some hung their heads, other gasped and covered their mouths with their hands; others laughed with embarrassment. In every case the children's behavior changed noticeably" (p. 147). Thus, neither the independent judges, nor the investigator, as biased as she might have been, mention stuttering as the outcome of this study. Additionally, the investigator makes strong statements in her conclusions concerning reduced verbal output, slower speaking rate, etc., but no data whatsoever were provided to back up such statements. From a scientific point of view, the absence of any reliability data renders interpretations of data, individual or group, as supporting treatment effects meaningless. Tudor's repeated remarks about the emotional and behavioral changes of the children are informative, but they do not constitute sufficient, credible evidence in support of her interpretations that they resulted from the experimental treatment. She was in no position to make meaningful pre-post treatment comparisons. She did not know the children prior to the study and failed to take elementary steps to obtain any baseline data about their personality, emotionality, and other behaviors to which she refers.

Assessments of the two types of quantifiable data, perceptual and speech, clearly indicate that all four experimental questions were answered in the negative. That is, the study failed to demonstrate any significant influence of labeling on the level of disfluent speech either in children who stutter or in children who do not stutter, under the experimental conditions as described. It is particularly apparent that the most critical experimental condition failed to show that more disfluency (or stuttering) was induced at the end of the study by means of occasional labeling in Group IIA, children classified as normally speaking. Consequently, the conclusion that the data supported Johnson's diagnosogenic theory in any way is untenable.

It is not clear when and why the speech-related difficulties reported for some members of Group IIA emerged. It is possible that an already existing mild stuttering was not detected, or was overlooked, at the beginning of the study, reinforced throughout the study, but still not detected at its conclusion. It is also possible, based on the investigator's comments, that the methods may have instilled negative reactions to speaking over the course of the study. It is less than likely, however, that stuttering would have emerged as a result of the treatment months after the treatment was terminated, when no stuttering was reported at the end of the study. It is very difficult, therefore, to know if, and how, the experimental procedures are related to the reported current long-term communication problems of some of the individuals from that group. Many factors that are completely independent of the study might have been operating, as Dryer reported other experts have also indicated.

Theory

In spite of the immense international popularity accorded to the diagnosogenic theory for more than three decades, dissenting voices were soon heard (e.g., Wingate, 1962). Four main types of scientific evidence have been mounted against it:

1. *The nature of speech disfluencies exhibited by children who begin stuttering and by normally speaking children.* A substantial body of data reported by several investigators has indicated that (a) although disfluencies in young children are normal, most children produce only a few of them (e.g., Yairi, 1981); (b) objective analyses show that disfluencies in the speech of children close to the time when they first begin stuttering are abnormal from the start of the disorder, and differ substantially, along several dimensions, from disfluencies of normally speaking children (Ambrose & Yairi, 1999; Throneburg & Yairi, 1994; Yairi & Lewis, 1984); Johnson et al.'s (1959) own disfluency data do not support his assertions as shown by McDearmon (1968), and (c) parents have often reported that they perceived abnormal speech in their children who stutter from the first day of stuttering onset (Yairi, 1983). Data-backed objections to the Johnson's theory were expressed by Yairi and Lewis (1984), stating that "The present findings, then, do not support the assertion that the disfluent speech behavior of children just regarded as stutterers is basically similar to the disfluent speech of those children not regarded as stutterers. Although other researchers, particularly Wingate (1962), objected to Johnson's diagnosogenic theory of the onset of stuttering, this study provides more direct evidence which calls the theory into question" (p. 154).

2. *Experimental punishment of stuttering.* Another type of evidence that negates the main assumptions of the theory is available from studies showing that adverse contingencies to stuttering, including negative verbal reactions (Cooper, Cady, & Robbins, 1970), loud sound bursts (Flanagan, Goldiamond, & Azrin, 1958), and electrical shock (Siegel and Martin, 1965) could result in an effect opposite to what would have been predicted by Johnson's theory; that is, substantial decline in stuttering. Experiments with preschool

children show the same effect. Martin, Kuhl, and Haroldson (1972) called attention to children's stuttering, yet stuttering dropped to near-zero levels.

3. *Parental correction of stuttering.* Studies in which parents report advising their children to "stop" and "slow down" when they stutter (obviously calling attention to it) showed that the children recovered from stuttering in spite of their parents' negative reactions (see review by Wingate, 1976). According to the diagnosogenic theory, these are exactly the reactions that should have increased stuttering.

4. *Non-environmental etiologies.* Strong evidence has emerged for non-environmental factors underlying stuttering etiology. Continuous indications during the past 40 years for strong genetic components to stuttering (Ambrose Cox, & Yairi, 1997; Howie, 1981; Kidd, Heimbuch, & Records, 1981) have provided impetus for linkage analyses studies design to identify general location of responsible genes. Indeed, Cox et al. (2000) reported three chromosomes suspected to contain loci for genes transmitting susceptibility to stuttering. In addition, using twin pairs where one or both twins stuttered, Felsenfeld et al. (2001) examined the proportions of genetic and environmental effects involved in the variability of the expression of stuttering. Results showed that about 70% can be accounted for by genetics, and about 30% are attributed to unique, or nonshared, environmental effects. There was no significant effect for shared environmental factors. Working in a different direction, other investigators have found anomalies in the speech-language area of the brains of people who stutter, that may be associated with increased risk for the development of stuttering (Foundas, Bollich, Corey, Hurley, & Heilman, 2001).

These and other developments of modern science have rendered the diagnosogenic theory obsolete as an acceptable explanation of the direct cause of stuttering. Based on current scientific evidence we reject the notion that several clinical sessions with Ms. Tudor, held at most on the average once per two weeks and complemented by ineffective orphanage staff participation, caused stuttering in most of the targeted children. Our re-analysis shows that, in fact, the Tudor study yielded the earliest evidence against the diagnosogenic theory. Had the study been published at the time and subjected to thorough scrutiny, it is quite possible that modern history of stuttering might have been quite different. One cannot deny, however, that the theory contains important elements. For example, the psychological dynamics that affect stuttering after it has begun, for which Johnson's theory may have had relevance, are not contested in our evaluation of the Tudor study. It is quite likely that parent-child interaction during the stage of stuttering plays an important role in the further course of the disorder. It is possible that the procedures employed in the Tudor study resulted in unpleasant, perhaps painful emotional reactions in the participants. Such influence, however, should not be confused with the cause of stuttering.

Ethics

The ethical issues that were raised in the *Mercury's* article regarding the Tudor study deserve a more thorough

analysis in several dimensions. First is the *use of human subjects*. Strong condemnations of the study and those responsible for it were expressed immediately after the newspaper article was published. In a follow-up article, the *Mercury News* (2001b) published a formal apology issued by the University of Iowa, and John Bernthal (2001), President of the American Speech-Hearing-Language Association, expressed reservations about the study. The present authors also stated that "it is unquestionable that the study was ethically wrong" (Yairi & Ambrose, 2001, p. 17). But the study must be viewed in the context of its time. To begin with, we believe that differences in standards that prevailed at that time should be recognized. Strict human subjects regulations did not exist and the research culture was considerably more lax than it is today. For example, it has been reported (Paden, 2001) that during that period, other investigators at the University of Iowa conducted laboratory experiments using gunshot noise (without live ammunition) to study the effect of startle on the stuttering of students. Thus, we are faced with the complicated question of whether ethics are relative or absolute. Specifically, when considering if the study was ethical, do the period in which it was conducted and that period's acceptable standards, context, and mindframe matter?

Inasmuch as there is willingness to recognize differences in standards that existed 60 years ago, the remaining major concern in the case of the Tudor study is whether or not the experimenter and her mentor intended to cause harm by turning normally speaking children into children who stutter. Our review of the study reveals no such apparent intent. The study investigated whether the level of disfluency could be changed as a result of labeling. It was *not* to create stutterers. Even if there was an unstated goal to increase disfluency to a level perceived as stuttered speech, there is no indication that Tudor or Johnson believed that, if successful, this would make the children chronic stutterers. This, in our opinion, is a critical point in judging the ethics of those involved in the conduct of the study. In this respect, one should also keep in mind that Johnson published a pamphlet in 1934 stating that the cause of stuttering is organic, primarily from interference with a child's natural handedness. He had only expressed the idea of negative evaluation of one's own disfluency as the first step towards stuttering in 1938, referring to listener reactions in a single sentence. Wingate's (2001) statement that Johnson had discussed labeling in these two sources is false. Thus, the Tudor study was completed three years before he briefly articulated his early ideas of the diagnosogenic theory in 1942. It is quite possible that his notions about the full relation between normal disfluency, listener reactions, and stuttering were still in a formative stage when the Tudor study was planned. The theory's full exposition came 20 years after the study was completed (Johnson et al., 1959), when he believed that his data gave credence to the theory.

It is also interesting to note that no criticism has been directed toward the part of the study that was designed to see if labeling would increase stuttering in children who were already stuttering. Is it more acceptable to make an

existing disorder worse? Indeed, there seems to be a tacit assumption that the increase is only temporary and, therefore, acceptable in lieu of the expected benefits to understanding the dynamics of the disorder. There is an appreciable body of scientific literature reporting studies designed to test conditions assumed to have the potential of temporarily increasing stuttering. Cooper, Cady, and Robbins (1970) tested effects of verbal reinforcement of disfluencies of stutterers and normally speaking controls. Oelschlaeger and Brutten (1975) paid stutterers for each time they stuttered to see if stuttering could be reinforced, as did Starkweather and Lucker with several children (1978). Other examples of operant conditioning experiments that attempted to increase stuttering are summarized in Costello and Ingham (1984). In fact, experimentation with delayed auditory feedback was thought to have elicited stuttering from normally fluent speakers (e.g., Lee, 1951). It appears to us that few would question the ethics of these experiments in their historical context. The point that we want to make is that Johnson may well have conceived the Tudor study within this frame of reference, assuming that any possible increase in disfluency would be temporary.

The setting of the study is also of concern. As previously mentioned, the study was carried out in an orphanage where parental consent was not required or even possible to obtain. Although this practice was common and sanctioned by public authorities, what is troubling is the failure of the investigators to disclose the true objectives of the study and their knowingly providing erroneous information to the orphanage's staff. The staff was told that Group IA (stuttering labeled as normal) really did not stutter, and that they should take the children's minds off their speech, and "don't, above all things, ever call it 'stuttering'" (p. 12). They were also told that children in Group IIA (normal speakers labeled as stuttering) actually were stuttering and that they should be monitored carefully, interrupting them, having them repeat, and not allowing "them to speak unless they can say it right" (p. 12). The staff were thus instructed to allegedly help these children.

A second dimension of research ethics of particular relevance to young investigators is that of *assigned responsibility*. According to present standards applicable to student research, both the investigator and the faculty mentor are responsible for carrying out research activities that conform with established regulations for the protection of human subjects, with the faculty mentor holding ultimate responsibility. Thus, inasmuch as someone is to be blamed, we question Tudor Jacobs' defense, as reported by the *Mercury*, that she conducted the study as instructed by Professor Johnson. This argument does not relieve her from responsibility. This is an important point both for educators and students engaged in scientific endeavors. Several war-related rulings (the Nuremberg courts after World War II, the Israeli court in the Eichman trial, and the courts of the United States Armed Forces in the My Lai trial), carry important lessons for less critical research situations. The courts ruled that obeying orders in committing unethical acts is not an acceptable defense. Certainly graduate students in a considerably more free

academic environment should be similarly responsible. Blame for the Tudor study (even if justified) cannot be entirely deflected to Professor Johnson. Regarding Dryer's assertion that there was a cover-up of the study by Johnson, it has, in fact, been available at the University of Iowa library and has been checked out by a good number of readers, including the present authors. Johnson may have ended collaboration with Tudor, and made no further reference to the study, but the actual document was available. Other theses he directed were also never published (e.g., Taylor, 1937).

A third dimension of ethics pertains to the *behavior and action of others parties* involved in reporting the Tudor Study. In this regard, the *Mercury's* decision to release personal information for nine identifiable participants in the Tudor study is questionable, whether or not the newspaper had permission to identify them by name and/or by number and age. Were the nine individuals aware that publishing their names or numbers allowed readers access to all of the personal information in the study, which is a public document, such as IQ scores? Also, while Dryer's decision to contact participants with the information that they were part of a study is legitimate and responsible, providing them with findings—of which the reporter's understanding and interpretation were clearly inadequate—is irresponsible. The participants have every right to know what really happened, but it is less than fair to provide them with inaccurate information.

Summary and Epilogue

A critical review of the Tudor study has revealed fundamental flaws in its design and execution. As it stands, the study failed to provide any credible scientific support that stuttering was produced "in the laboratory" or to lend other support to the diagnosogenic theory of stuttering. It appears, though, that the procedures employed caused unpleasant reactions in some of the children, which interfered with their verbal communication. Our assessment of the ethical issues suggests that the study should be viewed within the common standards of the period, that there is no evidence of intent to harm, and that the objective of increasing disfluent speech should not be confused with instilling chronic stuttering in normally fluent children. It is clear that such a project would not be allowed under present standards of the scientific community.

Finally, in spite of the controversy regarding the Tudor thesis, there is no question that Johnson's contribution to the study of stuttering remains very significant in many positive ways. Although his theory on stuttering onset has not prevailed, we must remember that sequences of errors are the stepping stones for progress in science. His work provided tremendous impetus for the study of various aspects of stuttering, particularly in calling our attention to the, then neglected, formative stages of the disorder. His clinical belief that if stuttering is learned, it can also be unlearned, has inspired many people who stutter to improve their speech.

On the positive side, Tudor Jacobs has expressed sincere regret for any harm that may have been caused by the study,

and Johnson's strong desire to promote the understanding and possible cure of stuttering was witnessed by many. His hopeful outlook is reflected in a letter to one of the present authors (Yairi) in which he wrote:

... you should give top priority to improving your own speech as much as you possibly can and you should aim at completely eliminating the behavior in which you indulge that interferes with your normal speech, the behavior which you call stuttering. It is, in my judgment, your own behavior and nothing more. On the basis of all the research I have done and the clinical observations that I have been able to make and the interpretations that I am capable of making of the information that I have been able to obtain, I would say that stuttering, so far as this word refers to something done by a speaker is concerned, consists of kinds of behavior, involving disruptive tensings, which interfere with what would otherwise be normal speech. You engage in this interfering behavior on the mistaken assumption, which you have learned, that if you do you will contend successfully with something that you call stuttering that will happen if you don't do these things. It is a false assumption. The stuttering is nothing but what you do trying to contend with the stuttering you expect to happen if you don't contend with it. To me, this is a very hopeful view of the problem. You can under the proper conditions and with the proper self instructions—the appropriate language of self-communication—learn to speak without doing the things you are now doing that interfere with your speaking (Johnson, 1963).

Let us not forget the Tudor study, but place it in proper perspective. Recent attention has brought to light many issues of ethics that have contemporary relevance, and the study is a powerful reminder that conclusions, even when consistently propagated, do not necessarily represent the quality or content of the original source. Misinterpretation is deceptively easy to come by, and can be difficult to counter. We do not have the luxury of time and resources to always re-read original sources that are not readily available. But when issues of importance arise, it must be done.

Acknowledgments

The preparation of this manuscript was partially supported by the National Institutes of Health, National Institute On Deafness and Other Communication Disorders, grant # 2 R01 DC00459, Principal Investigator: Ehud Yairi. The authors' opinions expressed in this manuscript do not reflect the position of NIH.

References

- Ambrose, N., Cox, N., & Yairi, E. (1997). The genetic basis of persistence and recovery in stuttering. *Journal of Speech and Hearing Research*, 40, 567–580.
- Ambrose, N., & Yairi, E. (1999). Normative data for early childhood stuttering. *Journal of Speech, Language, and Hearing Research*, 42, 895–909.
- Bernthal, J. (2001). Letters to the editor. Professional.asha.org announcements, June 13, American Speech-Language-Hearing Association.
- Bloodstein, O. (1995). *A handbook on stuttering* (5th ed). San Diego: Singular.
- Conture, E. (2001). *Stuttering: Its nature, diagnosis, and treatment*. Boston: Allyn & Bacon.
- Conture, E. & Kelly, E. (1991). Young stutterers' nonspeech behaviors during stuttering. *Journal of Speech and Hearing Research*, 34, 1041–1056.
- Cooper, E., Cady, B., & Robbins, C. (1970). The effect of the verbal stimulus words wrong, right, and tree on the disfluency rates of stutterers and nonstutterers. *Journal of Speech and Hearing Research*, 13, 239–244.
- Costello, J. & Ingham, R. (1984). Stuttering as an operant disorder. In R. F. Curlee & W. H. Perkins (Eds.), *Nature and treatment of stuttering: New directions* (pp. 187–313). San Diego: College Hill Press.
- Cox, N. (2000). Genetics of stuttering: Insights and recent advances. Paper presented at ASHA Convention, Washington, DC Abstract published in *ASHA Leader*, 5(16), 166.
- Dryer, J. (2001a, June 11). Theory improved treatment and understanding of stuttering. *The San Jose Mercury News*.
- Dryer, J. (2001b, June 13). Officials apologize for tests on stuttering. *The San Jose Mercury News*.
- Erard, M. (2001). Stutter, memory. *Lingua Franca*, 11.8, 40–45.
- Felsenfeld, S., Kirk, K., Zhu, G., Statham, C., Neale, M., & Martin, N. (2000). A study of the genetic and environmental etiology of stuttering in a selected twin sample. *Behavior Genetics*, 30, 359–366.
- Flanagan, B., Goldiamond, I., & Azrin, N. (1958). Operant stuttering: The control of stuttering behavior through response-contingent consequences. *Exper. Anal. Behav.*, 1, 173–177.
- Foundas, A., Bollich, A., Corey, D., Hurley, M., & Heilman, K. (2001). Anomalous anatomy of speech-language areas in adults with persistent developmental stuttering. *Neurology*, 57, 207–215.
- Howie, P. (1981). Concordance for stuttering in monozygotic and dizygotic twin pairs containing stutterers. *Journal of Speech and Hearing Disorders*, 24, 317–321.
- Johnson, W. (1934). Stuttering in the preschool child. *Child Welfare Pamphlets No. 37*, University of Iowa: Iowa City.
- Johnson, W. (1938). The role of evaluation in stuttering behavior. *Journal of Speech and Hearing Disorders*, 3, 85–89.
- Johnson, W. (1963). Personal communication to Ehud Yairi.
- Johnson, W., Van Riper, C., Davis, D., Scarbrough, H., Hunsley, Y., Bakes, F., Travis, L., & Dwyer, S. (1942). A study of the onset and development of stuttering. *Journal of Speech Disorders*, 7, 251–257.
- Johnson, W., and Associates. (1959). *The onset of stuttering*. Minneapolis: University of Minnesota Press.
- Kidd, K., Heimbuch, R., & Records, M. (1981). Vertical transmission of susceptibility to stuttering with sex-modified expression. *Proceedings of the National Academy of Science USA*, 78, 606–610.
- Korzybski, A. (1941). *Science and sanity: An introduction to Non-Aristotelian systems and general semantics* (2nd ed). New York: International Non-Aristotelian Library Publishing Company.
- Lee, B. (1951). Artificial stutter. *Journal of Speech and Hearing Disorders*, 16, 53–55.
- Martin, R., Kuhl, P., & Haroldson, S. (1972). An experimental treatment with two preschool stuttering children. *Journal of Speech and Hearing Research*, 15, 743–752.
- McDearmon, J. (1968). Primary stuttering at the onset of stuttering: A reexamination of data. *Journal of Speech and Hearing Research*, 11, 631–637.

- Oelschlaeger, M., & Brutten, G.** (1975). Response-contingent positive stimulation of the part-word repetitions displayed by four stutterers. *Journal of Fluency Disorders*, 2, 10–17.
- Onslow, M., Costa, L., Rue, S.** (1990). Direct early intervention with stuttering: Some preliminary data. *Journal of Speech and Hearing Disorders*, 55, 405–416.
- Otto, F., & Yairi, E.** (1976). A disfluency analysis of Down's syndrome and normal subjects. *Journal of Fluency Disorders*, 1, 26–32.
- Paden, E.** (2001). Personal communication to Ambrose and Yairi. (Aus: month & day if possible?)
- Ratner, N.** (2001). Coordinator's corner. *Special Interest Division 4, Fluency and Fluency Disorders*, 11.3, 1–4.
- Siegel, G., & Martin, R.** (1965). Experimental modification of disfluencies in normal speakers. *Journal of Speech, Language, and Hearing Research*, 8, 235–244.
- Starkweather, W., & Lucker, J.** (1978). Tokens for stuttering. *Journal of Fluency Disorders*, 3, 167–180.
- Throneburg, R., & Yairi, E.** (1994). Temporal dynamics of repetitions of during the early stage of childhood stuttering: An acoustic study. *Journal of Speech, Language, and Hearing Research*, 37, 1067–1075.
- Taylor, G.** (1937). *An observational study of the nature of stuttering at its onset*. Unpublished master's thesis, University of Iowa.
- Tudor, M.** (1939). *An experimental study of the effect of evaluative labeling on speech fluency*. Unpublished master's thesis, University of Iowa.
- Wingate, M.** (1962). Evaluation and stuttering: III. Identification of stuttering and the use of a label. *Journal of Speech and Hearing Disorders*, 27, 368–377.
- Wingate, M.** (1976). *Stuttering theory and treatment*. Boston: Irvington Publishers.
- Wingate, M.** (2001). Failure didn't modify Johnson's views. Letter, *ASHA Leader* 6.22, 35.
- Yairi, E.** (1981). Disfluencies of normally speaking two-year-old children. *Journal of Speech, Language, and Hearing Research*, 24, 490–495.
- Yairi, E.** (1983). The onset of stuttering in two- and three-year-old children. *Journal of Speech and Hearing Disorders*, 48, 171–177.
- Yairi, E.** (1997). Disfluency characteristics of childhood stuttering. In R. Curlee and G. Siegel (Eds.), *Nature and treatment of stuttering* (pp. 49–78). Needham Heights, MA: Allyn and Bacon.
- Yairi, E., & Ambrose, N.** (2001). The Tudor experiment and Wendell Johnson: Science and ethics reexamined. *ASHA Leader*, 6.13, 17.
- Yairi, E., & Lewis, B.** (1984). Disfluency near the onset of stuttering. *Journal of Speech, Language, and Hearing Research*, 27, 154–159.
- Yaruss, S.** (1997). Clinical implications of situational variability in preschool children who stutter. *Journal of Fluency Disorders*, 22, 187–204.

Received August 31, 2001

Accepted February 11, 2002

DOI: 10.1044/1058-0360(2002/018)

Contact Author: Nicoline Ambrose, University of Illinois, Speech and Hearing Science, 901 S. Sixth Street, Champaign, IL 61820. E-mail: nambrose@uiuc.edu