REPORT OF THE RUTGERS RESEARCH ADVISORY BOARD

Investigation into Allegations of Research Misconduct Against Dr. William Brown

April 25, 2012

I. HISTORICAL BACKGROUND

Drs. Trivers, Palestis and Zaatari (2009), in a paper titled "Anatomy of a Fraud," accused Dr. William Brown of committing research misconduct and including false research results in Brown et al. (2005), a paper published in *Nature* with Dr. Trivers as a co-author. Dr. Brown and another coauthor of this same *Nature* paper (Dr. Cronk), have denied these charges in a written rebuttal (Brown and Cronk, 2009). The research in question was funded by the National Science Foundation (NSF) which was acknowledged in the paper.

Rutgers University became aware of these accusations in 2009 and, following NSF guidelines and University policy, completed an inquiry into these allegations. The recommendation from this inquiry, as noted on December 22, 2009 in a letter from Dr. Pazzani to NSF, was to undertake a full investigation of the allegations. The NSF agreed with this recommendation and, on February 18, 2010, asked Rutgers to undertake the full investigation.

We begin this report of the full investigation with a summary of our findings and then present a more detailed description of the study design, the charges of misconduct, and the actions we took to investigate the charges. We conclude the report with detailed explanations of our analysis of the evidence for the three main allegations that were made against Dr. Brown by Drs. Trivers, Palestis and Zaatari (2009) and the reasoning for the findings.

II. SUMMARY OF OUR FINDINGS –- WE FIND THAT:

- Substantial (clear and convincing) evidence exists that research fraud has occurred in several areas. Evidence exists that:
    - Based upon the investigator's (Dr. Brown's) knowledge of subject performance, or access to existing evaluations of subject performance, there was biased selection of subjects who were to be included in the symmetry / asymmetry comparison groups so as to artificially obtain desired results;
    - There was falsification of averaged data scores from the Jamaican children's evaluations;
    - There were omissions of data availability and documentation, as well as conflicting data sets that are consistent with a cover-up against the charges.

- The study design is very complicated and, in some instances, not well defined by the investigators. There are multiple copies of files, some with > 80,000 data fields, for analysis. There are innumerable accusations and rebuttals.
    - The scale and complexity of the study make it very hard for us to document each and every allegation against Dr. Brown that was made by Drs. Trivers, Palestis and Zaatari (2009) in such a way that:
        - Will be easily understood even by analytically oriented persons;
        - Will address all conceivable rebuttals that could be made by the accused and accusers.
    - With these concerns in mind, we decided to focus on the most substantive of the allegations.
        - This report makes no findings with regard to whether other allegations regarding Dr. Brown's research are well founded or not.

III. BRIEF SYNOPSIS OF THE STUDY IN QUESTION (Brown et al., 2005)

Hypothesis – Persons (Jamaican children) who are more physically symmetrical will be perceived to be better dancers by their peers. This will be true more so for males than for females.

Data Source - Ongoing Study of ~183 Jamaican Children that began in the middle 1990s

- Measures of Symmetry were taken on the Jamaican Children in 1996 and 2002:
    - Summed relative absolute differences between left and right side of body were used to calculate what is defined as "fluctuating asymmetry" (FA).
    - This was a cumulative score of mean adjusted absolute differences in relative asymmetry size (FA) summed across nine body parts, with a higher score indicating a person is more asymmetric.

        NOTE – While we believe that the score more accurately should be labeled Relative Fluctuating Asymmetry (RFA), due to the adjustment of the differences by the mean score, the term FA is used in the paper so we also do so here in the report to be consistent.

    - The calculations of fluctuating asymmetry and summed score are described in more detail later in this report.

- Videotapes of dancing were made on some of the children in 2004/2005.
    - As part of a complicated process described later, the same Jamaican children in the main study were also later the judges of the dancing ability of the selected 40 Jamaican dancers described below.

Study Analysis Sample chosen in Early 2005

- From the larger subject population, 40 children (20 girls and 20 boys) were selected into four groups based on the following stated criteria:

Asymmetric boys = 10 boys who were in the "top 1/3rd" for FA asymmetry scores both in 1996 and again in 2002

Symmetric boys = 10 boys who were in the "lowest 1/3rd" for FA asymmetry scores both in 1996 and again in 2002

Asymmetric girls = 10 girls who were in the "top 1/3rd" for FA asymmetry scores both in 1996 and again in 2002

Symmetric girls = 10 girls who were in the "lowest 1/3rd" for FA asymmetry scores both in 1996 and again in 2002

NOTES

1. While there were ~183 original subjects eligible to be selected for the 40 dancers, the number of potential subjects was smaller, ~106 due to various reasons including missing FA data in 1996 and/or 2002, and/or not later being filmed dancing.

2. It is not totally clear from the paper (Brown et al., 2005) whether the top/bottom 1/3rd means for both boys and girls pooled together or if the top/bottom 1/3rds were calculated for each sex separately; the indications were that the sexes were pooled together. Nor is it clear from the paper whether the pools of subjects used for consideration in 1996 and 2002 included all persons with available data for the given year of consideration or were restricted to persons with complete data for both 1996 and 2002.

3. From the allegations made by Drs. Trivers, Palestis and Zaatari (2009) and the rebuttal of Drs. Brown and Cronk (2009), more than 10 subjects were eligible to be included in three of the above four groups and the process of how the total available pool of subjects were reduced to 10 for these three groups is one of the salient issues in the allegations of misconduct.

4. There are also claims in the rebuttal of Drs. Brown and Cronk (2009) that persons with what was deemed to be "poor dancing tape quality" were excluded from consideration for these four symmetry/asymmetry groups. But Dr. Brown and Dr. Cronk presented no evidence as to how this was done nor any information as to whether and how such exclusions were documented.

5. Importantly for the charges of fraud being made, before these 40 subjects were chosen, two Rutgers undergraduates had evaluated the dancing tapes and it appears that these scores were available to Dr. Brown prior to the selection process. Drs. Trivers, Palestis and Zaatari (2009) claim this on page 13. In Drs. Brown and Cronk's rebuttal (2009) it states that the Rutgers undergraduate evaluations of the tapes were "not all available" at the time the 40 dancers were selected. But Drs. Brown and Cronk do not elaborate further on this to indicate what portion of the Rutgers undergraduate evaluations were available at that time. However, it is clear on page 3 of the rebuttal by Drs. Brown and Cronk (2009) that the dance animations had been viewed by the Rutgers undergraduates and that at least some of the dance scores assigned by these undergraduates were used as part of a grant application made on February 23, 2005 by Dr. Brown, which was prior to the selection of subjects to groups. Thus it is not disputed that Dr. Brown had access to at least some of the Rutgers undergraduate evaluations of the dance animations before the selection of the 40 dancers. In any case, none of this elaborate prescreening of the subjects is

mentioned in the *Nature* (2005) paper or in any available Appendices to the paper.

Evaluation of Dancing Outcome in March 2005

1. The same Jamaican Children (155 of the 183 children) evaluated the "digitalized" dance routines.

    - Digitalized means that the identity and appearance of the dancer was hidden.

2. The plan was for each of the 155 children to evaluate each of the tapes from the digitalized dancers.

    - The overall score of each of the 40 tapes was the average score of all children who evaluated the tapes with these caveats:

        i. If a child evaluated his/her own digitalized dance, that score was excluded;
        ii. There appear in the data (c.f. sent to us by Dr. Palestis as described later in this report) evaluations that were incomplete or incorrect and may have thus been excluded from the mean score. But this is not fully documented in the article or by other information sent to us by Dr. Brown (or by Dr. Palestis).

IV. SUMMARY OF THE ALLEGATIONS OF MISCONDUCT THAT WERE MADE BY DRS. TRIVERS, PALESTIS AND ZAATARI (2009)

1. Parties Involved:

    a. The party alleged to have engaged in research misconduct:

        - Dr. William Brown – First author of Brown et al. (2005) who was a post-doctoral student in Anthropology at Rutgers when this work was done. He is now a faculty member at The University of Bedfordshire in Bedford, UK. He was one of 7 authors on the paper.

    b. The parties alleging research misconduct (Drs. Trivers, Palestis, Zaatari, 2009):

        - Dr. Robert Trivers – Senior author of the paper Brown et al. (2005) and in the same Department as Dr. Cronk (Anthropology).

- Dr. Brian Palestis – Not a coauthor of Brown et al. (2005) and has not been affiliated with Rutgers (he is at Wagner College). He apparently has done the statistical analysis to "confirm" and support Dr. Trivers' position.
- Dr. Darine Zaatari – Not a coauthor of Brown et al. (2005). She was a Ph.D. student at Rutgers when the *Nature* paper was written and has since graduated. She was apparently involved in much of the initial investigation by the accusers.

c. Dr. Lee Cronk – The second author of Brown et al. (2005) and the Principle Investigator on the Grant. He is not accused of committing any misconduct. but has come to the defense of Dr. Brown.

2. Sequence of accusations and rebuttals *as stated by the accusers and then the accused* (paraphrasing the words used).

   a. Soon after the publication of Brown et al. (2005), Dr. Trivers (through communications with persons who were unable to obtain the same results that were in the paper from what they believed to be the data used for the analyses) developed concerns about the data Dr. Brown had used and analyses that appeared in the article. He and the other accusers began contacting Dr. Brown for explanations and the specific data sets that Dr. Brown used. (A series of emails resulting from these contacts was sent to us by Dr. Palestis).

   b. Not being satisfied with Dr. Brown's response, Drs. Trivers, Palestis and Zaatari (2009) conducted their own analysis of the data and facts as they saw them, ultimately leading to the point where they concluded that fraud had been committed by Dr. Brown.

   c. Drs. Trivers, Palestis and Zaatari attempted to have the journal "*Nature*" publish a letter retracting the article. When *Nature* refused to do this, they attempted to have an expose published in another journal. When this did not happen, they published their own 91 page analysis "*Anatomy of a Fraud*" (Trivers, Palestis and Zaatari, 2009).

   d. When contacted by the Rutgers Office of the General Council about the document "*Anatomy of a Fraud*," Drs. Brown and Cronk prepared a ~50 page (including appendices) rebuttal to the accusations (Brown and Cronk, 2009).

3. The Major Accusations by Drs. Trivers, Palestis and Zaatari:

   a. Dr. Brown falsified some of the 1996 and 2002 fluctuating asymmetry (FA) scores on <u>selected subjects</u> in a fashion that i) moved boy/girl dancers to whom the two Rutgers undergraduates had given worse dance ratings into the top 1/3rds of the FA asymmetry scales (most asymmetric) for 1996 and 2002

(and thus caused these worse dancers to meet the selection criteria for being asymmetric) and ii) moved boy/girl dancers who had been accorded better dance ratings by the two Rutgers undergraduate students into the bottom 1/3rds of the FA asymmetry scales for 1996 and 2002 (and thus caused these better dancers to meet the selection criteria for being symmetric).

- The hypothesis, as presented in the allegations, was that Jamaican children and the Rutgers undergraduates would rate the dancers in about the same way. The specific allegation is that Dr. Brown leveraged this possibility to spike the top 1/3^{rds}asymmetric group with bad dancers and the bottom 1/3rdsymmetric group with good dancers.

b. More than 10 boys or girls met the criteria to potentially be included in three of the four groups. When that happened, Dr. Brown selected the 10 who would be included into the groups. The allegation is that he did this in a biased fashion so as to selectively choose from the eligible subjects those who were rated as worse dancers by the Rutgers undergraduates to place into the asymmetric groups and similarly selectively chose subjects who were rated as better dancers by the Rutgers undergraduates into the symmetric groups.

c. After the Jamaican children dance evaluations were collected and scored, Dr. Brown falsified the Jamaican children's dancing score ratings to enable results which statistically supported the hypothesis of the paper.

d. NOTE – As was stated earlier in this report, Drs. Trivers, Palestis and Zaatari's 2009 document further contains a multitude of other accusations. We did not investigation every allegation, but focused on those where it could be efficiently proved and substantively determined that misconduct occurred. (c.f. by Brown) in relation to the 2005 *Nature* paper.

V. SUMMARY OF THE ACTIONS WE HAVE TAKEN TO INDEPENDENTLY INVESTIGATE THE CHARGES AND REBUTTALS

1. The Rutgers Office of the General Council had already requested and received data sets relevant to the accusations when this committee became involved. We nevertheless requested from both Dr. Brown and Dr. Palestis (who did much of the analysis for the Trivers, Palestis and Zaatari (2009) report) all data sets and materials that had relevance to the allegations and, in particular, instructions on how to calculate a) the 1996 and 2002 FA scores and b) the Jamaican student dance rating scores from the raw data.

   - Dr. Palestis responded within 1 week of the request and sent us, among other things, data sets that included:

i. The data for 1996 and 2002 asymmetry scores which Dr. Palestis said Dr. Brown had sent him earlier and data for the same scores that exists in the database that Dr. Trivers' group maintains for the ongoing Jamaican study. (Dr. Brown did not collect the FA score data himself but received it from others who had collected these data).
ii. The individual ~155 Jamaican students' dance ratings of the 40 study subjects. Dr. Palestis said he had received these data earlier from Dr. Brown.
iii. Instructions on how to calculate all scores in the data described above in i. and ii.
iv. Summaries of Dr. Palestis', Trivers' and Zaatari's comparative analyses of the data they received from Dr. Brown with their own data from the Trivers group database and the results reported in the 2005 *Nature* article.

- Dr. Brown responded later:

i. He did send his data for the 1996 and 2002 asymmetry scores which after making a considerable number of comparisons appear to us to be essentially the same data that Dr. Palestis had sent us which he stated he had received from Dr. Brown.
ii. Dr. Brown indicated that he no longer had raw data on the Jamaican students' ratings of the 40 dancers. When we questioned him further about these data, Dr. Brown said that the data Dr. Palestis et al. used for their analysis of dance scores (i.e. attributed to being from Dr. Brown) must be old or corrupt and the correct data could not be recovered from it. Quoting Dr. Brown from his January 25, 2011 email to Dr. Pazzani, "*I sent a file to Dr. Brian Palestis some time ago, but it appears that this file is either corrupted or an earlier version of the one used by the research assistant (i.e., to decide which ratings would be included in the average). ….. If I could find the file or figure out how to calculate the averages from the one I sent Dr. Brian Palestis, I would send it to you along with detailed instructions to help with the investigation. …. Nonetheless, I will look for the file I sent to Dr. Palestis and attempt again to reconstruct the average ratings."* We have not received any further correspondence from Dr. Brown on this issue.

2. We requested from all participants who we believed might have access to these documents (Drs. Trivers, Brown and Cronk) to send us copies of earlier versions of the paper and the reviews, most notably the original paper that was submitted to *Nature* along with the review and the response to the review.

- The first response was from Dr. Cronk who sent us multiple copies of earlier versions of the paper.

- Dr. Brown followed with one copy of an earlier version of the paper.

- Dr. Trivers sent a copy of the review report of the first submission of the paper.

3. To be thorough, we sent emails and/or made phone calls to other coauthors (Drs. Keith Grochow, Amy Jacobson, C. Karen Liu, and Zoran Popovic) asking if they had any knowledge that could be relevant to the investigation. None of these authors were alleged to have engaged in misconduct, and while some were mentioned in the rebuttal, it did not seem as if they would have knowledge pertinent to the charges.

    - Dr. Popovic responded (and we believe he was also speaking for Drs. Liu and Grochow) that they had been aware of these allegations for quite some time, had been contacted about them by several sources and, as coauthors of the paper, were anxious to know the findings of the investigation. They had no significant new information on this to share with us.

    - Dr. Jacobson responded that she had no role in the study beyond being the field site manager for the general project and thus could not help us further.

4. After we had undertaken our analysis and were ready to finalize the report, we sent emails to Dr. Brown and Dr. Cronk requesting explanations on two findings we made that could reflect inconsistencies or fraud in the study design and analysis.

    - Dr. Cronk met with us at Rutgers in early October 2011 and Dr. Brown sent a written reply in December 2011.

VI. SUMMARY OF OUR ANALYSIS AND FINDINGS

We first followed the approach that Drs. Trivers, Palestis and Zaatari (2009) had used (including examining the rebuttals made by Drs. Brown and Cronk (2009)) to see if the arguments put forth were valid and then to see if we could replicate the different analyses with the data sets we had been given. While we found that the previous approaches which had been used were well reasoned and exhaustive, we tried to streamline and distill the analyses to be more easily understandable, communicable and addressable. The result is the following findings relevant to the main allegations of fraud that were made in the last paragraph of page 5 of Trivers, Palestis and Zaatari (2009).

1. *Allegation* - The 1996 and 2002 FA asymmetry scores of the 40 dancers who were chosen for the study groups were systematically fabricated in a fashion to make better dancers more symmetric and worse dancers less symmetric.

    Our Conclusion – There is clear and convincing evidence to support the allegations that this alleged research misconduct occurred.

a. This fabrication occurred and did cause dancers who were rated better by the Rutgers undergraduates to be more likely to be inserted into the "symmetric" boys and girls groups and dancers who were rated worse by the Rutgers undergraduates to be more likely to be inserted into the "asymmetric" boys and girls groups.

b. It does not seem possible that; this fabrication i) could have happened by chance, ii) could have been perpetrated by anyone other than Dr. Brown or iii) if had it been perpetrated by someone other than Dr. Brown, that Dr. Brown would not have noticed this problem and reported it after years of questioning by Dr. Trivers' group and then by us.

2. *Allegation* - When Dr. Brown had the opportunity of choosing 10 subjects from a group of more than 10 to make the final top/bottom symmetry group for boys and girls, he chose the subjects in a way that favored the alternative hypothesis (i.e. based on the Rutgers undergraduate students dance evaluations).

   Our Conclusion - There is clear and convincing statistical evidence to support the allegations that the alleged research misconduct occurred. Dr. Brown either used the data collected by the Rutgers undergraduates (or some other informed evaluations of the digitalized dances) to carefully select subjects as alleged. Thus, for three of the four groups, among eligible dancers, those with better Rutgers undergraduate ratings were placed into the symmetric groups and those with poorer Rutgers undergraduate ratings were placed into the asymmetric groups.

3. *Allegation* - Dr. Brown fabricated the Jamaican children averaged dance score summaries of the 40 dancers in order to obtain statistically significant findings that supported the alternative hypothesis.

   Our Conclusion – There is enough evidence to support that the alleged research misconduct occurred. Dr. Brown is unable to produce data that can support the findings he reported in *Nature* (2005) which, as both the first author and as the person who undertook that data analysis, he should be able to do. However, Dr. Palestis produced a data set he claims to have received from Dr. Brown. Dr. Brown subsequently acknowledged he sent this data to Dr. Palestis and sent the same data to us, but claims that this data is incorrect / unusable and that he no longer has the correct data. It is thus impossible to know exactly what was done in the analysis by Dr. Brown because he relies on claims of unwritten / undocumented or otherwise unexaminable reasons for exclusions and/or incorrectness of some values in this existing data. Nonetheless, the findings of our analyses on the only existing raw dancer rating data initially provided by Dr. Palestis, are very consistent with those of Trivers et al. and are incompatible with the findings reported in *Nature* (2005).

We now present detailed explanations of our findings on the three main allegations of research misconduct that were made against Dr. Brown.

1. ***Allegation*** **- The 1996 and 2002 FA asymmetry scores of the 40 dancers who were chosen for the study groups were systematically fabricated in a fashion to make better dancers more symmetric and worse dancers less symmetric.**

   Our Conclusion **- There is clear and convincing evidence to support the allegations that this alleged research misconduct occurred.**

   a. **This fabrication occurred and did cause dancers who were rated better by the Rutgers undergraduates to be more likely to be inserted into the "symmetric" boys and girls groups and dancers who were rated worse by the Rutgers undergraduates to be more likely to be inserted into the "asymmetric" boys and girls groups.**

   b. **It seems impossible that; i) this fabrication could have happened by chance, ii) it could have been done by anyone other than Dr. Brown, or iii) had it been by someone other than Dr. Brown, that Dr. Brown would not have noticed this problem and reported it after years of questioning by Dr. Trivers' group and then by us.**

EVIDENCE

<u>For Fabrication of Asymmetry Scores</u>

A. The 1996 and 2002 asymmetry scores in the data sets sent to us by Dr. Palestis were entirely internally self-consistent (i.e. the data did not contradict itself) with respect to the Fluctuating Asymmetry (FA) scores and their component variables.

B. The 1996 and 2002 FA scores and their components in the data sent to us by Dr. Brown were:
   1) Internally self-consistent for all subjects who were not chosen to be one of the 40 dancers.
   2) In general, not internally self-consistent (data contradicted itself) for the 40 subjects who were chosen to be dancers as described in Section C below.

C. The non-self-consistency of FA scores in Dr. Brown's data is, in our view, impossible to explain by anything other than fabrication of some of the data by a person who, at the time of the fabrication, did not realize that the other items also needed to be changed for the data to be self-consistent or otherwise did not think to change these items.

For each subject, the Fluctuating Asymmetry (FA) score was calculated as a sum of absolute relative asymmetry for 9 body parts (elbow, wrist, knee, ankle, foot, ear, 3rd digit, 4th digit and 5th digit) as described below.

FA = $\sum_{P} RA_P$ where P = 1,…, 9 enumerates the nine body parts and RA_P is the relative asymmetry of the given body part (i.e. hand, ear, 4th digit, etc.)

For each body part, $RA_P = \frac{AD_P}{M_P}$ with

AD_P = Absolute Value of [Left Side Measure – Right Side Measure]

M_P = Average of Left Side Measure and Right Side Measure

The values of AD_P, M_P and RA_P are saved in the data sets we received from Drs. Brown and Palestis for each person and body part P (Dr. Brown's data sent to us is missing AD_P for the 3rd digit in 1996 and for the ears in 2002). As described above for each subject and body part, if we go into the data sets and for the Pth body part and year (1996, 2002) take the ratio of the values AD_P / M_P of a given child, this ratio is always the same as the value of RA_P for that Pth body part of that child during the same year in the data (i.e. self-consistent) in Dr. Palestis' data as it should be. The observed ratio AD_P / M_P is also always equal to (i.e to within three decimal places) the value of RA_P for the same body part of the child in the same year in Dr. Brown's data *for all subjects not selected into the study*; with any differences that were less than 3 decimal places being very small (i.e. of order < 10^{-10}) and thus being likely due to round off error at some stage and otherwise having no impact on the FA score.

However, the ratios of AD_P / M_P are largely not equal (within 3 decimal places) to the values of RA_P for the same body part of the same child in the given year (1996 or 2002) *for almost all of the 40 subjects selected to the study groups* for Dr. Brown's data among the body parts that were included in the FA score.

For example, with P = 4thdigit, going to subject 15 in 1996 (who was selected as one of the 40 dancers) in Dr. Brown's data we observe

AD_P = 0.875

M_P = 55.888

RA_P = 0.0076 (which rounded to three decimal places in Tables 1 and 2 is 0.008)

But looking at AD_P / M_P for this person gives the self-consistent value of RA_P as 0 .875 / 55.888 = 0.0157 (which rounded to three decimal places in Tables 1 and 2 is 0.016)

In their rebuttal to the allegations by Drs. Trivers, Palestis and Zaatari (2009), Drs. Brown and Cronk (2009) suggest that some data discrepancies might be due to "rounding" errors. However, it is obvious that the difference between 0.0157 and 0.0076 is too large to be due to round-off error and that this difference does not qualitatively change by only taking the measures of AD_P and M_P out to 3 Vs.4 or more decimal places. The same is true for the other inconsistencies we observed between RA_P and AD_P / M_P in Dr. Brown's data in the 40 selected dancers.

Furthermore, it should be noted that for all study subjects and all body parts P, the values of AD_P and M_P for body part P of any given subject do not differ between Dr. Brown's and Dr.

Palestis' data sets. The inconsistent values of RA_P do not equal the ratios of the corresponding AD_P / M_P for the vast majority of body parts in the 40 selected dancers in Dr. Brown's data set and differ from the RA_P in Dr. Palestis' data set (which always equals the ratio of the corresponding AD_P / M_P). As we just noted, *when these differences between Dr. Brown's and Dr. Palestis' data occur*, the RA_P in Dr. Palestis' data is equal to the ratio of the corresponding AD_P / M_P while the RA_P in Dr. Brown's data does not equal the ratio of the corresponding AD_P / M_P .

In other words, if we look at the 4th digit of subject 15 for 1996 in Dr. Palestis' data, we see the correct and self-consistent values

AD_P = 0.875

M_P = 55.888

RA_P =0.0157 (i.e. = 0.875 / 55.888)

Due to there being 9 body parts measured on 2 different years (1996 and 2002) and 290 subjects in the data set, we cannot show all the comparisons here. However, Table 1 displays the values of AD_P, M_P , the actual ratio of these values AD_P / M_P and RA_P for the 4th digit in 1996 among the first 30 dancers in Dr. Brown's data set which includes some who were selected into the 40 asymmetric / symmetric dancers. Those who were selected into the final 40 asymmetric / symmetric dancers are highlighted in red in Table 1. When the recorded RA_P does not equal the ratio of the corresponding AD_P / M_P, the last 2 columns in Table 1 are highlighted in bold. Table 2 shows the same comparisons for the 40 selected dancers. For 34 of these subjects, the RA_P does not equal the corresponding ratio AD_P / M_P. As is true for the other subjects and body parts in 1996 and 2002, the recorded RA_P always equals the observed ratio AD_P / M_P in subjects who were not selected to be in the 40 dancers but usually does not for those who were selected.

It should be noted that measures for AD_P M_P and RA_P are recorded for 1996 in Dr. Brown's dataset on one body part (the hand) that was not used in the FA score. For this body part, the ratio of AD_P / M_P always equals the corresponding RA_P in the 40 selected dancers in spite of the fact just noted above that it seldom does for the body parts that were included in the 1996 FA score. It should also be noted that sometimes values for AD_P and M_P are present but the corresponding result for RA_P is missing in Dr. Brown's data. For example, this happens with ID 7 in Table 1 and ID 287 in Table 2. However, we have found that in settings when this happens an entire set of values for at least one other summed body part in that year is missing. For example, ID 7 is missing measures of AD_P M_P and RA_P for foot in 1996 and ID 287 is missing the values of AD_P M_P and RA_P for elbow in 1996. Thus the missing RA_P'*s* for the 4th digit of IDs 7 and 287 in 1996 could reflect that all of the RA_P'*s* needed for the 1996 sum were not available for those IDs. (While Dr. Brown in fact has a sum recorded for 1996 FA of ID 287

as shown in column 2 of Table 3 that is mentioned later in this report, this was not possible as column 3 of the same table shows since elbow was missing.)

When we met with Dr. Cronk in October 2011, he had no explanation for the discrepancies between the recorded RA_P and the actual ratios AD_P / M_P for body parts among the 40 selected subjects. Dr. Brown also acknowledged the inconsistencies existed as well when he replied to our questions on December 1, 2011, but his only explanations alluded to the fact that either he did not know how they could happen and/or that the data we had received from him may not have been the same data that he actually used in 2005 and/or that these errors may have been introduced by other people before he received the data.

To quote (with salient phrases underlined by us) from part of his response to our questions on this issue that he returned on December 1, 2011 …. *"This is interesting as you rightly point out the hand was not used as one of the FA's in the composite. Recall that all information that was used and presented in the Nature paper was not from the master dataset I sent you. Any values that are included in this file were pasted from the file used in 2005. This is clear evidence that the file I was working with in 2005 is indeed different from the file you attached as I previously claimed".* **……..** *It is challenging to explain why these inconsistencies occur. Recall that when making the dataset for Dr Palestis well after the dance paper was published in Nature (the email and file time stamps indicate this fact) it came to my awareness that there were errors. I should point out that these initial errors were introduced before I began working on the project. Indeed to make the so-called master file for Dr Palestis involved me merging, cutting and pasting from different files some of which I no longer have access. Since errors were discovered after I made the file I am skeptical about the validity of this file. You have discovered another problem, to which I have no logical explanation. I acknowledge it to be there, but as to how it emerged (and when) is unclear to me. Without the original files I was working with it difficult to isolate how and when discrepancies emerged in this post-publication dataset."*

The only explanation we can see for the non-self-consistencies in Dr. Brown's data is that Dr. Palestis' data set is correct and that the values for RA_P were altered in Br. Brown's data so that they would sum to the values of FA for those subjects in 1996 and 2002 which had also been altered. But this was only done within the 40 selected dancers and was done by someone who was either not aware that the corresponding values for AD_P and M_P also needed to be altered to make the data self-consistent or otherwise did not bother to do so. We see no conceivable way this alteration could happen by chance or accident; we conclude it must be the result of fabrication. For example, non-self-consistencies between AD_P and M_P were RA_P observed at least once in 39 of the 40 selected dancers compared to never in the 66 other filmed dancers with available FA data for 1996 and 2002 who were not selected. The P-value for this to occur by chance alone is less than one in 10^{-27} times by exact test.

The Alteration of Asymmetry Scores Was Done by Dr. Brown

It seems impossible that anyone else except Dr. Brown (who did the data analysis for the paper and held the data set) would have access to these data to alter only the values of RA_p and corresponding summed FA_p. We do not see how someone creating a data set in 2005 before Dr. Brown began working on the project would have the reason or ability to alter these values only among those 40 people who ultimately at a later date became selected to be dancers using what is now an incompletely defined process and, what would have been at the time of that alteration, an unknowable process.

The Alteration of Asymmetry Scores Favored the Investigator's Hypothesis in a Way That Could Have Been Foreseen by Dr. Brown

The complexity of the study design and fact that this design was not clearly explained (and further confused by caveats such as persons were excluded from consideration because their videos were deemed un-evaluable) complicates a certain determination of "what would have happened" if the data had not been fabricated as we believe it was. However, we compare in Tables 3 and 4 respectively the differences [Dr. Brown's data summed FA – Correct Summed FA] for 1996 and 2002 respectively. By "Correct Summed FA" we mean the summed FA that is self-consistent with the AD_P and M_P in the data set. For example, in Table 3 for ID 15, the value for summed 1996 FA in Dr. Brown's data was 0.110 (in column 2). However, based on the actual values of AD_P and M_P for the 9 body parts in 1996 and their ratios, the correct (i.e. self-consistent) 1996 FA for ID 15 was 0.163 (in column 3). This means that Dr. Brown's summed 1996 FA for ID 15 was shifted -0.053 (in column 4) from the correct value (-0.053 = 0.110 – 0.163) making that person more symmetric than they would be by the self-consistent FA measure. Column 5 has the averaged Rutgers undergraduate dancer scores for ID 15 which was 123.93. Now 123.93 was one of the higher scores meaning this person's summed FA was shifted lower by 0.053 to make this person more symmetric by Dr. Brown's score, and this person was also rated as a relatively good dancer by the Rutgers undergraduate students. The format for Table 4 is the same as that for Table 3 except that 2002 rather than 1996 FA scores are involved.

In order to see if the shifts (from self-consistent) in the 1996 and 2002 FA scores in Dr. Brown's data were associated with the Rutgers undergraduates' dance scores, we examined the correlations of the shifts (column 4) with the averaged Rutgers undergraduate scores (Column 5) in Tables 3 and 4 among those dancers where Dr. Brown's value differed from the self-consistent value. These analyses were restricted to only those subjects in 1996 and 2002 respectively, where Dr. Brown's FA differed from the correct self-consistent FA. For 1996 (Table 3) the shift between Dr. Brown's value and the self-consistent value was negatively correlated with the averaged Rutgers undergraduate dancer scores (ρ= -0.39 with P=0.0157 for no association by Formula 16.25 in Berenson and Levine, 1999). For 2002 (Table 4) the shift between Dr. Brown's value and the self-consistent value was also negatively correlated with the averaged Rutgers undergraduate dancer scores ρ= -0.24 for 2002 with P=0.245, (by Formula 16.25 in Berenson and Levine, 1999). This means that, compared to bad dancers, good dancers were more shifted towards symmetry by the alterations in Dr. Brown's FA scores in both 1996 and 2002, something that would support the alternative hypothesis.

Using Fisher's (1950) method (as described below on page 23 of this report) to pool the p-values from 1996 and 2002 together with the fact that the shifts were in the same direction gives an overall two-sided P-value of 0.0152 for the shifts in 1996 and 2002 simultaneously being directionally associated with Rutgers undergraduate scores. In other words, it is not likely that the shifts in the FA scores that Dr. Brown's data had from the correct self-consistent FA scores for 1996 and 2002 would correlate with the averaged Rutgers undergraduate evaluations in the direction of the alternative hypothesis as strongly as they did.

It should be noted that Drs. Brown and Cronk's rebuttal (2009) claims that some or all of Rutgers undergraduate evaluations were not available when the 40 symmetric / asymmetric dancers were selected. But even if that were the case, it does not invalidate the findings of this test which indicate that the changes in FA within Dr. Brown's data were directionally associated with a supposedly independent measure of the dancing ability. For example, others (including we believe almost certainly Dr. Brown) were also able to view the animation tapes before the 40 dancers were selected. Thus Dr. Brown could have used dancer evaluation information from sources other than the Rutgers undergraduate students to base any decisions for fabrication. As these dancer evaluations from other sources would also likely agree with the Rutgers undergraduate students with respect to quality of dance, the fabricated shifts in FA would still be statistically associated with the Rutgers undergraduate scores in the direction of the alternative hypothesis even if the undergraduate scores were not used in the fabrication process. The rebuttal from Drs. Brown and Cronk (2009) mentions tapes being excluded from consideration for selection by the investigators due to poor quality, an assertion that means that the tapes must have been viewed in advance to screen for this. It stands to reason that the perception of Dr. Brown and others on dancing ability would be in the same directions as that of the Rutgers undergraduate students and, if so, this association of shifts in FA from the self-consistent value to Dr. Brown's value with Rutgers undergraduate ratings would transfer to the same associations with other ratings of dancing ability as well.

2. ***Allegation*** **- When Dr. Brown had the opportunity of choosing 10 subjects from a group of more than 10 to make the final top/bottom symmetry group for boys and girls, he chose the subjects in a way that favored the alternative hypothesis (i.e. based on the Rutgers undergraduate students dance evaluations).**

   Our conclusion **- There is clear and convincing statistical evidence to support the allegations that the alleged research misconduct occurred. Dr. Brown either used the data collected by the Rutgers undergraduates or some other informed evaluations of the digitalized dances to carefully select subjects as alleged. Thus, for three of the four groups, among eligible dancers, those with better Rutgers undergraduate ratings were placed into the symmetric groups and those with worse Rutgers undergraduate ratings were placed into the asymmetric groups.**

EVIDENCE

With respect to this charge (that there was a biased pre-selection of the 10 subjects when more than 10 were eligible such that those chosen were biased in the direction of the alternative hypothesis when the Jamaican students evaluated the tapes), the background may be summarized as follows: 167 individuals were assessed for FA in 1996 and 2002. Of these, according to Trivers, Palestis, Zaatari (2009), 167 were filmed while dancing using a motion capture technique of whom 106 had complete FA data for 1996 and 2002. It was then decided that the effect of FA on perceived dance ability would be compared across four groups of 10 individuals each: symmetrical males, asymmetrical males, symmetrical females and asymmetrical females. To identify the 10 subjects for each group that would be drawn from the larger population, a criterion was established, namely that each of the 10 subjects for each group must fall in either the i) the upper thirds of the symmetry-asymmetry scale for both 1996 and 2002 or ii) the lower thirds of the symmetry-asymmetry scale for both 1996 and 2002. Dr. Brown's review of his FA data for both years using these criteria identified 13 "symmetrical" eligible males, 13 asymmetrical eligible males, 10 symmetrical eligible females and 16 asymmetrical eligible females (Trivers, Palestis, Zaatari 2009; Brown and Cronk 2009). That is, for three of the four groups, there were too many possible subjects and 10 subjects needed to be selected from the pool. The charge against Dr. Brown is that the selection process was not random or blind but done deliberately with the intent of increasing the probability that the main alternative hypothesis would be statistically substantiated.

The 40 dance animations were ultimately evaluated by 155 Jamaicans who had also served as dancers or dancer candidates to provide the outcome data for Dr. Brown's study. However, as noted earlier, the animations were pre-evaluated by two undergraduate dance students of Rutgers University. Dr. Brown allegedly had access to these evaluations and, allegedly, used them to select the 40 animations from the larger pool of eligible subjects as described above. However, even if Dr. Brown did not have access to these Rutgers undergraduates' dance evaluation scores, he and/or others had access to the tapes and their own ratings of these tapes might be similar to

those of the Rutgers undergraduate students. So, is there evidence either way that Dr. Brown did or did not use a randomized / blind procedure to select the 40 subjects from the larger pool of 52 = 13 + 10+ 13 + 16?

The explanations from the Trivers, et al. and Dr. Brown as to how Dr. Brown proceeded to select (or eliminate) subjects differ from each other. Significantly, the *Nature* paper (Brown et al. 2005) does not mention that this was done let alone how it was done. But in this quote from page 47 of *Anatomy of a Fraud*, Drs. Trivers, Palestis and Zaatari (2009) attribute Dr. Brown as saying*:*

"First I randomized subject numbers for the entire data set using web-based software (www.random.org). Afterwards, random selection was done through a roll of the dice. Specifically if 14 males were in the top third percentile for time one (1996) and time two (2002) a dancer was eliminated if my dice rolled a "one" for any one of those 14 males."

But later in 2009, Drs. Brown and Cronk responded to the charges of Drs. Trivers, Palestis and Zaatari in a slightly different manner (page 4).

"Selection of the forty animations was a three-step process. First, in order to make the process blind with respect to subjects' ID numbers, an online random number generator was used to generate a temporary ID number for each one. Second, the temporary ID numbers were put into a hat, drawn by a research assistant, and jotted down by Brown. This was intended to randomize not only the selection of the animations but also the order in which they were shown. Finally, to reduce each category to the required ten animations, Brown rolls a die. If he rolled a one for a particular animation, then that animation's randomly assigned temporary ID was taken out of consideration."

A few points are worthy of note:

1) The first mention of this elimination process (see page 48 of Trivers, Palestis, Zaatari (2009)) is apparently in a Dec 20, 2005 e-mail from Dr. Brown to Dr. Palmer who had asked Dr. Brown several questions about the methodology used in the Brown et al., (2005) paper. We do not have this e-mail but it follows the Brown et al. (2005) publication. Nowhere in the *Nature* publication is there any mention of the fact that the dance animations were already evaluated by the Rutgers undergraduates prior to being included in the study or that Dr. Brown already knew (or had the opportunity to know) the dance ability of the 40 dancers he selected. Likewise, there is no mention in *Nature* (2005) of the elaborate process he claims he used to make sure his selection of the 10 dance animations, when more than 10 were eligible, was both blind and randomized.

2) The selection process is extraordinary in its complexity and lack of documentation given that all Dr. Brown had to do was use a random number generator to generate random numbers for all

eligible subjects in a group and pick those 10 subjects with the largest (or smallest) numbers, and the program used could have been easily saved and documented.

3) In their rebuttal to this claim, Brown and Cronk (2009) indicated that some of the eligible dancers may not have been selected because of flaws in the animations. But none of this is mentioned in the *Nature* paper nor are there any references to such invalidations having been done in the data sets we received from Dr. Brown. Furthermore, the Brown and Cronk rebuttal also noted that there was considerable demonstrated heterogeneity in opinions between different observers as to what constitutes a flaw and which dance animations were flawed. This leaves open the explanation that flaws could be found or otherwise claimed as a means of excluding observations that were not desired and thus negates the usefulness of this explanation.

Drs. Trivers, Palestis and Zaatari (2009) presented two different mathematical "proofs" of their contention that Dr. Brown non-randomly selected his subjects. We find that one of these approaches was successfully invalidated by Dr. Brown and Cronk's (2009) rebuttal. To that end, we looked at this supposition independently using what we think was the best mathematical model, which turned out to be an extension of the approach used by Drs. Trivers, Palestis and Zaatari (2009) in Table 2 of their critique. Trivers, Palestis and Zaatari (2009) indicated that there were 13, 13, 9 and 16 potential candidate subjects in the study groups of interest: symmetrical males, asymmetrical males, symmetrical females and asymmetrical females, respectively by the eligibility criteria of being (based on Dr. Brown's values of summed FA) in the highest $1/3^{rd}$ of summed FA for both 1996 and 2002 or lowest $1/3^{rd}$ of summed FA for both 1996 and 2002. Dr. Brown had included one ineligible subject to get 10 symmetrical females. Only 14 of the 16 candidates to be asymmetrical females had undergraduate Rutgers dance evaluations. Drs. Trivers, Palestis and Zaatari (2009) also provided the subject IDs of these eligible subjects.

There is incomplete information about how the top and bottom $1/3^{rds}$ of symmetry were defined. That is, was it: i) based on all subjects pooled together or based on girls and boys separated, ii) based on all subjects with FA data in a given year or restricted to subjects with complete data in 1996 and 2002 and/or iii) excluding subjects deemed to have poor quality videos? Therefore, it is impossible for us to be sure of and to replicate the exact analysis Dr. Brown used to identify candidate subjects in each group. However, Drs. Brown and Cronk (2009) did not rebut the Trivers, Palestis, Zaatari (2009) claim as to the candidate subjects in their reply but, rather, sought to show that the 10 selected for inclusion into each group were not done so in a statistically biased fashion. We therefore focus here on the same subjects claimed by Drs. Trivers, Palestis and Zaatari.

Figure 1 presents in separate blocks: the 13, 13, and 14 candidate subjects for symmetrical males, asymmetrical males, and asymmetrical females that had Rutgers undergraduate evaluations of their dance scores. The IDs within each block are sorted on the averaged Rutgers undergraduate dance evaluation score from lowest to highest. The IDs that were not selected to be in the final

10 dancers are highlighted in red. To the right of each block is an arrow that points in the direction of selection (in terms of Rutgers undergraduate dancing scores) that would support the alternative hypothesis. For example, the left most block is symmetrical males. The Rutgers undergraduate dance scores for the 13 candidate members in this group ranges from 110.75 to 138.8. The direction of selection which supports the alternative hypothesis of symmetrical persons being better dancers is for those with higher dance scores to be included in the 10 members of this group. The three IDs not selected were 178, 189 and 70. Looking at all three groups one can see that the selected dancers (in black) tend to aggregate in the direction of Rutgers undergraduate scores that supported the alternative hypothesis.

We now present what we believe to be the best approach to statistically test whether the selection process in the final 10 subjects for these 3 groups favors the alternative hypothesis. The Table below presents the means and standard errors we obtained of Rutgers undergraduate student scores for those selected and those not selected to be in the group based on our calculations (which are essentially similar to those from Drs. Trivers, Palestis and Zaatari (2009)).

Category	Selected Dancers Rutgers Undergrad Ratings			Eligible non-Selected Dancers Rutgers Undergrad Ratings			P-Values, t-test (two sided)*
	Mean	Std-Err	N	Mean	Std-Err	N	
Sym Boys	122.80	2.34	10	116.10	2.19	3	0.185
Asym Boys	94.03	3.36	10	118.04	3.03	3	0.0035
Sym Girls	N/A			N/A			
Asym Girls	97.94	5.30	10	122.87	4.55	4	0.053

*P-values two sided. Using Mann-Whitney rank test gives similar results with P-values of 0.12, 0.007 and 0.014 respectively.
N/A – Only 10 subjects in this group (including 1 added that was not eligible) so all were selected.

For the symmetrical boys group, the mean of the Rutgers undergraduate dance rating was higher for those selected than those not selected (P=0.185, two sided t-test) which is in the direction of the hypothesized difference. For the asymmetrical boys and girls each, the mean Rutgers undergraduate dance rating was lower for those selected than for those not selected in each group (P=0,.0035 and P=0.053 by two sided t-tests, respectively) which again is in the direction of the hypothesized difference.

To simultaneously quantify the statistical significance of the deviations in all three symmetry-sex groups according to both their strength and direction with respect to the alternative, we used Fisher's (1950) method to pool the P-values from the individual tests of these groups. Namely

that under the null hypothesis $-2\sum_{i=1}^{3}\log_e(p_i^*/2)$ will have a χ^2 distribution with 2 * 3 = 6 degrees of freedom where p_i^* is the two-sided P-value and for this setting so $p_i^*/2$ always gives the one sided P-value in the direction of the alternative. Based on the two-sided P-values of 0.185, 0.0035 and 0.053 and the direction always favoring the alternative, the one-sided p_i^* obtained from dividing these by 2 results in $\chi^2 = 24.68$ which has a P-value of 0.00039 which we multiplied by 2 to get 0.00078 to convert it to a two-sided hypothesis which would also allow for directional selection opposing the alternative hypothesis. In other words, the two-sided chance that Dr. Brown would simultaneously choose 10 dancers each from the 13, 13, and 14 eligible subjects that favored the alternative hypothesis in terms of the Rutgers undergraduate dance scores by chance alone is only about 8 in 10,000. Similar results are obtained if Fisher's (1950) method is applied to Mann-Whitney rank test p-values rather than t-tests to compare selected and non-selected subjects.

The allegation by Dr. Trivers, Palestis and Zaatari (2009) suggests that Dr. Brown first fabricated the FA scores and then, once the groups of eligible subjects were created, selected the 10 subjects in each group. We believe another possibility could be that the 10 subjects desired in each group were first selected with disregard to (or lack of knowledge of) the summed FA scores. Then once this was done, the summed FA scores were fabricated in order to make these subjects eligible for selection. Both scenarios are consistent with the finding that there was little chance that the 10 subjects obtained in each group from the eligibles would so strongly support the alternative hypothesis in terms of the Rutgers averaged undergraduate dance scores.

3. ***Allegation*** **- Dr. Brown fabricated the Jamaican children averaged dance score summaries of the 40 dancers in order to obtain statistically significant findings that supported the alternative hypothesis.**

   Our Conclusion **– There is enough evidence to support the allegations that the alleged research misconduct occurred. Dr. Brown is unable to produce raw dancer rating data that can support the findings he reported in *Nature* (2005) which, as both the first author and as the person who undertook that data analysis, he should be able to do. However, Dr. Palestis produced a data set he claims to have received from Dr. Brown. Dr. Brown subsequently acknowledged he sent this data to Dr. Palestis and sent the same data to us, but claims that this data is incorrect / unusable and that he no longer has the correct data. It is thus impossible to know exactly what was done in the analysis by Dr. Brown because he makes claims of unwritten / undocumented or otherwise unknowable reasons for exclusions and/or incorrectness of some values in this existing data. Nonetheless, the findings of our analyses on the only existing raw dancer rating data initially provided by Dr. Palestis, are very consistent with those of Trivers et al. and are incompatible with the findings reported in *Nature* (2005).**

EVIDENCE

Dr. Brown did not initially provide us a raw data set of the individual Jamaican children's evaluations of the dance scores and effectively stated that the summarized averages of the Jamaican children's evaluations of these dancers were all he had. Dr. Brown's email to us on October 14, 2010 stated, …… *"I do not have the original raw files of the dance ratings made by each rater, just the final averages used in the Nature paper. As per Professor Trivers' instructions the research assistant used just the bad and good dancer rating item (there were other dance quality items as well) to calculate the average, only if it was 50 percent consistent with itself across the other dance quality items for a particular dancer and all the other dancers viewed by that rater. Professor Trivers' assumed that this would save money and that if a rater was variable in their ratings then perhaps they did not take the rating task seriously (or did not understand the task). I should point out that there were other constraints in calculating the dance composite scores (e.g., self-evaluations were removed, incorrect sex detections removed)."*

Dr. Palestis, however, did provide us a raw data set of the individual Jamaican children's evaluations of the dance scores which he stated was sent to him from Dr. Brown. Later after being specifically asked to send all relevant raw data that were used to calculate the average Jamaican rater dance scores for the 40 dancer in a January 21, 2011 email from Dr. Pazzani, Dr. Brown (on February 7th 2011) sent us this same data with an explanation that this data must be corrupt or otherwise impossible to use. To quote Dr. Brown's email on January 25, 2011 apparently describing this dataset, *"I sent a file to Dr Brian Palestis some time ago, but it appears that this file is either corrupted or an earlier version of the one used by the research assistant (i.e., to decide which ratings would be included in the average). If I could find the file or figure out how to calculate the averages from the one I sent Dr Brian Palestis, I would send it to you along with detailed instructions to help with the investigation."* And Dr. Brown goes on in that email to imply that we ask Dr. Trivers to send us the original rating sheets for the dancers

and apparently use the detailed instructions that are with those sheets to reconstruct the scores ourselves.

To further evaluate this issue, we analyzed the only data set of individual rater scores available which was originally sent to us by Dr. Palestis, to see if there was some way we could obtain the same overall dancer ratings from this data that Dr. Brown reported in the *Nature* (2005) paper. Following instructions that were sent to us by Dr. Palestis on how to analyze the data (as no instructions were sent by Dr. Brown beyond that it was probably futile to try to analyze it), we proceeded to check that self-evaluation results were eliminated from the data, but had to subjectively eliminate other values such as 0's which were out of place suggesting that no valid evaluation was done. However, the instructions themselves were self-evident from the data. For example, 1) from left to right the columns contained data from dancer IDs sorted in the numerical order assigned by the study and 2) from the evaluation columns, the cell had no data by design when the evaluator row and dancer column were the same person who was not allowed to evaluate his/her own dance. Then for each of the 40 dancers in columns, we calculated the overall rating of that dancer as the mean of all ratings among the eligible subset of 155 rows (raters) who evaluated that dance. For example, if 131 of the 155 raters had eligibly evaluated a given column-dancer, then the overall score of that column-dancer was the mean of those 131 row-evaluations.

We tried different ways of applying what Drs. Trivers, Palestis and Zaatari (2009) stated possibly could be "filter variables" in the raw dancer ratings data set to rule out what could have been deemed to be inferior dance ratings. For example, if 50 of the 131 row-raters of a given dancer had a value of "0" for the variable in the column to the right of the ratings column and "0" meant to not use, then the overall score would be the mean of the remaining 81 (= 131 – 50) row-raters. However, no matter how we interpreted these "potential filter variables," in no case did we even come close to replicating the average scores that Dr. Brown reported. Table 5 compares the average scores that appear in the summarized dancer rating data Dr. Brown sent (Column 2) to those we obtained (Column 3) from the individual raters as well as averaged scores that Dr. Palestis sent to us obtained by two other methods (i.e. of subjectively eliminating values that seemed to indicate no valid rating was done) that his group reported that they had used ("Darine D" in Column 4 and "Mean Dan" in Column 5). For example, for ID 15, Dr. Brown reports an average score of 48.90. We obtained 63.43 as did the "Darine D" approach in Dr. Palestis' analysis. The "Mean Dan" approach of Dr. Palestis obtained an average of 62.20. Our values for the averages were generally very close to the "Darine D" and "Mean Dan" values sent by Dr. Palestis (in fact often being identical to the Darine D values). Among the four sets of values, those by Dr. Brown were almost always the outlier, deviating greatly from the other three. Again we tried using variables that Drs. Trivers, Palestis and Zaatari (2009) thought could be filter variables to eliminate bad dance ratings. But as they had also reported, we were not able to obtain results even close to those of Dr. Brown's in any of these attempts.

In the Table below are i) the means and standard deviations of the averaged Jamaican children's ratings of the 4 dance groups in dancer score using the dancer-average scores we calculated among all raters for the dancer in the raw data initially sent by Dr. Palestis, followed by ii) the means and standard deviations of these scores in the four dance groups that Dr. Brown reported in the revised paper to *Nature* which was published, followed by iii) the means and standard

deviations of the Jamaican children's ratings of the 4 dance groups in dancer score we directly calculated using the summarized values in the summarized data Dr. Brown first sent us.

Means and Standard Deviations of Averaged Dance Scores by Symmetry Sex Group

Method / Report	Symmetrical Males	Asymmetrical Males	Symmetrical Females	Asymmetrical Females
i) Our Analysis From raw Data initially sent by Dr. Palestis	58.24(±**16.12**)	40.95(±**15.80**)	46.71(±**19.73**)	37.93(±**15.13**)
ii) The Paper Published in Nature	57.31(±**10.65**)	39.22(±**9.23**)	45.53(±**9.47**)	35.58(±**9.70**)
iii) Our Analysis of Means in Dr. Brown's Data	57.31(±**10.65**)	39.22(±**9.23**)	45.53(±**9.47**)	35.58(±**9.70**)

Our analyses of the "already averaged" dancer ratings in the data set that Dr. Brown first sent in iii) up above produced identical results to those that were reported in the Nature paper in ii) up above. However, our analyses of the rater averaged values we obtained from averaging the ratings of individual evaluations for each of the 40 dancers in the raw data initially sent by Dr. Palestis in i) above produced different group means and dramatically smaller within group standard deviations.

We next followed the same approach that Trivers, Palestis and Zaatari (2009) had used to see if the differences in the summarized dancer scores that we had obtained would result in materially different findings than those that Dr. Brown had reported were used for the *Nature* (2005) paper. As Trivers, Palestis, and Zaatari (2009) had concluded before, we also concluded that the findings of symmetry effect and interaction with gender that had been very significant in the *Nature* (2005) paper were either not statistically significant or were barely statistically significant in the analyses of our summarized scores in the raw data initially sent by Dr. Palestis. For example, using a t-test to compare i) the two-sided test for a difference in asymmetrical and symmetrical boys using the summarized values we obtained from the individual dancer ratings in the raw data initially sent by Dr. Palestis would be $t = 2.42$ $P= 0.03$ (marginally< 0.05) compared to $t = 4.06$ $P <0.001$ from the "already averaged" values that Dr. Brown reported, ii) the two-sided test for comparing asymmetrical and symmetrical girls using the summarized values we obtained from the individual dancer ratings in the raw data initially sent by Dr. Palestis would be $t = 1.12$ $P=0.58$, not statistically significant compared to $t = 2.32$ $P= 0.03$ from the "already averaged" dancer ratings in the file sent by Dr. Brown, iii) the gender-symmetry interaction term would also not be significant, two-sided $p= 0.32$ from ANOVA compared to $p=0.04$ that was

reported in the *Nature* article. Again, Drs. Trivers, Palestis and Zaatari (2009) reached similar conclusions using the overall ratings of the individual dancers in the file with the approaches they used to obtain these averaged means. We strongly believe, as did Trivers, Palestis and Zaatari (2009), that *Nature* would not have published a paper with associations as weak as those we observed or that Trivers et al.(2009) had observed based on the individual ratings of the dancers in the field they purported to have received from Dr. Brown.

We noticed as did Trivers, Palestis and Zaatari (2009), that the standard deviations within each of the four study groups that we had obtained analyzing "Dr. Brown's averages" were only half as large as those when analyzing the averages we obtained from the raw data (as well as when analyzing the Darine D and Mean Dan averages that Trivers, Palestis and Zaatari (2009) had obtained). We then noticed that there was a large drop in the group standard deviations occurring between the initial version of the paper submitted to *Nature* and the revised version. The statistical reviewer of the initial submission had noted that the original methods of analysis were incorrect and the authors (Dr. Brown and colleagues) had agreed with these comments. We asked both Dr. Cronk and Dr. Brown to explain the reasons for the change in group means and in particular the large drop in group standard deviation between the original submission to *Nature* and revision and for Dr. Brown to demonstrate it and/or provide details. While Dr. Brown did not fully explain the original method of analysis that was used, his explanations did include enough details for us to see how such a drop could, in theory, happen with a change in analysis along the lines that Dr. Brown suggested had happened. Further exploration of this issue here will not be productive as the only raw data available was the file initially provided by Dr. Palestis (and then provided by Dr. Brown) that Dr. Brown is disputing as valid. It has already been shown that analysis of this raw data produces different results from those reported by Dr. Brown.

Finally, it should be noted here that one could also argue that more complicated two-way designs that simultaneously adjust for reviewer effects while looking at dancer effects nested within symmetry-gender group could have been fit to the data, and this would be a better approach. But follow up on this would not bring any bearing on the issue of fraud as these other approaches had not been used by Dr. Brown. Applying an analytical approach that was not used in the original analysis (even if this approach is more correct) to data that Dr. Brown has stated is incorrect will not yield any more answers than what has already been gained.

Our analysis of the individual dancer ratings of the 40 dancers initially provided by Dr. Palestis and then from Dr. Brown, confirms the findings of Trivers, Palestis and Zaatari (2009) that: a) for the 40 dancers, the overall averaged rater dancer means differ dramatically from those present in the summarized data provided by Dr. Brown, and b) these differences caused the sex/symmetry group means to differ and within group standard deviations to be lower in a way that produced much more statistically significant findings against the null hypotheses in the data Dr. Brown reported. Again, we do not believe that the paper would have been accepted into

Nature with the less statistically significant results obtained from the data we received from Dr. Palestis.

As Dr. Brown was the person who performed the analysis on the study, we believe that he has the responsibility to keep copies of all final data and analyses (including a correct data set of the raw dancer ratings) for an extended period of multiple years, if not indefinitely. This is not difficult to do with current technology. We do not find it credible that i) Dr. Brown would lose and not be able to recover the correct raw data while retaining old /corrupt raw data if he had known he did nothing wrong, especially as the response from Drs. Brown and Cronk (2009) indicates that at least Dr. Cronk was aware of Dr. Trivers' concerns about the validity of the analysis as far back as Spring 2006 shortly after the analysis published in *Nature* had been conducted; and ii) Dr. Brown is not able to provide specific details as to what is wrong with the existing raw dancer rating data set that he and Dr. Palestis provided that are actionable for correction. We therefore believe that the preponderance of the evidence supports the allegation that Dr. Brown falsified the summarized dancer scores used in the final analysis for the *Nature* (2005) paper.

REFERENCES

1. Berenson ML, Levine DL. Basic Business Statistics. Prentice-Hall Upper Saddle River, New Jersey 1999.

2. Brown WM, Cronk L, Grochow K, Jacobson A, Liu CK, Popovic Z, Trivers Dance Reveals symmetry especially in young men *Nature* 438:22;1148-1150, 2005.

3. Brown WM, Cronk L No Fraud: A response to Trivers et al. Letter to Rutgers General Council November 10, 2009.

4. Fisher RA. Statistical Methods for Research Workers. Oliver and Boyd, London 1950

5. Trivers R, Palestis BG, Zaatari D. The Anatomy of a Fraud, Symmetry and Dance. TPZ Publishers Antioch CA, 2009

Table 1 - Comparison of the Actual Ratios AD_P / M_P to the RA_P Recorded in Dr. Brown's Data for the 4th Digit in 1996 among first 30 Dancers in Data (Those Selected into final 40 Dancers in Red)

Dancer ID	1996 Fourth Digit Absolute FA(AD_P) in Dr. Brown's Dataset[a]	1996 Mean Size Fourth Digit (M_P) in Dr. Brown's Dataset[b]	True Ratio AD_P / M_P	1996 Value of Rel FA (RA_P) for Fourth Digit in Dr. Brown's Dataset[c]
1	2.075	53.863	0.039	0.039
2	0.550	64.425	0.009	0.009
3	0.675	64.938	0.010	0.010
4	0.600	62.600	0.010	0.010
5	1.025	62.663	0.016	0.016
6	0.400	63.075	0.006	0.006
7	2.175	60.988	0.036	Missing
8	1.575	59.988	0.026	0.026
9	0.210	61.045	0.003	0.003
10	0.575	58.488	0.010	0.010
11	0.325	61.338	0.005	0.005
12	1.275	58.588	0.022	0.022
13	0.975	48.513	0.020	0.020
14	0.100	60.675	0.002	0.002
15	0.875	55.888	**0.016**	**0.008**
16	3.300	66.525	0.050	0.050
17	1.100	55.700	0.020	0.020
18	0.350	60.825	0.006	0.006
19	1.650	55.800	0.030	0.030
20	1.625	57.038	0.028	0.028
21	1.100	57.450	**0.019**	**0.031**
22	1.025	58.638	0.017	0.017
23	1.500	60.450	0.025	0.025
24	0.150	58.175	0.003	0.003
25	1.250	60.900	0.021	0.021
26	0.425	62.488	0.007	0.007
27	1.675	56.413	0.030	0.030
28	1.175	59.038	0.020	0.020
29.	2.225	57.913	0.038	0.038
30	1.250	64.450	**0.019**	**0.000**

a. From Column R of Master_File_2006_Data_Brian_Excel_Version(1).xls
b. From Column S ofMaster_File_2006_Data_Brian_Excel_Version(1).xls
c. From Column T of Master_File_2006_Data_Brian_Excel_Version(1).xls

NOTE – All values are rounded to three decimal places

Table 2 - Comparison of the Actual Ratios AD_P / M_P to the RA_P Recorded in Dr. Brown's Data for the 4th Digit in 1996 among the 40 subjects selected to be Dancers

Dancer ID	1996 Fourth Digit Absolute FA(AD_P) in Dr. Brown's Dataset[a]	1996 Mean Size Fourth Digit (M_P) in Dr. Brown's Dataset[b]	True Ratio AD_P / M_P	1996 Value of Rel FA (RA_P) for Fourth Digit in Dr. Brown's Dataset[c]
15	0.875	55.888	**0.016**	**0.008**
21	1.100	57.450	**0.019**	**0.031**
23	1.500	60.450	0.025	0.025
30	1.250	64.450	**0.019**	**0.000**
33	1.350	52.125	**0.026**	**0.037**
34	0.650	59.300	**0.011**	**0.022**
38	1.350	63.650	**0.021**	**0.014**
55	0.825	56.813	0.015	0.015
63	3.800	62.975	0.060	0.060
67	0.025	58.013	**0.000**	**0.017**
68	0.725	60.463	**0.012**	**0.011**
75	3.100	54.500	**0.057**	**0.068**
86	2.400	55.425	**0.043**	**0.019**
89	1.750	59.875	**0.029**	**0.017**
94	0.050	52.250	**0.001**	**0.011**
103	0.900	55.000	**0.016**	**0.006**
110	1.100	59.625	**0.018**	**0.030**
113	0.475	64.313	**0.007**	**0.017**
115	0.675	55.588	**0.012**	**0.027**
117	3.025	52.963	**0.057**	**0.052**
119	0.700	59.700	**0.012**	**0.034**
139	0.250	51.600	**0.005**	**0.016**
152	0.200	55.875	**0.004**	**0.002**
162	0.025	64.763	**0.000**	**0.002**
175	1.400	69.550	**0.020**	**0.019**
182	0.350	73.300	**0.005**	**0.000**
185	0.300	56.725	0.005	0.005
192	3.550	62.750	**0.057**	**0.073**
194	1.450	61.725	**0.023**	**0.025**
195	0.875	62.763	**0.014**	**0.025**
197	0.975	61.488	**0.016**	**0.015**
200	0.925	57.238	0.016	0.016
203	0.475	57.838	**0.008**	**0.005**
205	4.550	53.950	**0.084**	**0.092**
206	1.675	61.088	**0.027**	**0.042**
222	1.050	54.950	**0.019**	**0.028**
229	2.025	55.013	**0.037**	**0.031**
235	0.375	63.788	**0.006**	**0.017**
239	0.225	62.088	**0.004**	**0.002**
287	0.375	51.388	0.007	Missing

a. From Column R of Master_File_2006_Data_Brian_Excel_Version(1).xls
b. From Column S ofMaster_File_2006_Data_Brian_Excel_Version(1).xls

c. From Column T of Master_File_2006_Data_Brian_Excel_Version(1).xls

NOTE – All values are rounded to three decimal places

Table 3 – Differences Between Dr. Brown's RA_P and Correct (i.e. Self-Consistent with data set AD_P / M_P ratio) Summed FA in 1996 and Rutgers Undergrad Dance Evaluations Among 40 Selected Dancers

ID	Brown1996 FA	Correct 1996 FA	Brown – Correct 1996 FA	Averaged Rutgers Undergraduate Student Ratings
15	0.110	0.163	-0.053	123.93
21	0.254	0.144	0.110	98.15
23	0.122	0.121	0.000	138.80
30	0.126	0.241	-0.115	100.58
33	0.285	0.185	0.100	87.275
34	0.246	0.146	0.100	89.45
38	0.130	0.178	-0.048	109.80
55	0.105	0.098	0.007	129.575
63	0.211	0.211	Same	107.575
67	0.269	0.124	0.145	100.725
68	0.090	0.099	-0.009	135.50
75	0.239	0.139	0.100	73.93
86	0.102	0.206	-0.104	104.48
89	0.134	0.216	-0.082	118.98
94	0.206	0.116	0.090	105.80
103	0.269	0.345	-0.075	75.45
110	0.247	0.147	0.100	112.98
113	0.218	0.128	0.090	109.48
115	0.287	0.157	0.130	95.28
117	0.135	0.171	-0.036	110.75
119	0.285	0.085	0.200	91.25
139	0.217	0.117	0.100	87.50
152	0.101	0.114	-0.013	121.40
162	0.082	0.067	0.015	119.55
175	0.224	0.236	-0.011	67.88
182	0.085	0.158	-0.073	121.38
185	0.092	0.096	-0.005	123.13
192	0.284	0.134	0.150	99.38
194	0.126	0.114	0.013	102.55
195	0.240	0.140	0.100	113.775
197	0.105	0.110	-0.004	120.50
200	0.115	0.114	0.001	115.50
203	0.087	0.111	-0.024	127.40
205	0.236	0.169	0.067	108.35
206	0.288	0.158	0.130	82.60
222	0.262	0.182	0.080	99.40
229	0.109	0.157	-0.048	121.13
235	0.269	0.169	0.100	113.53
239	0.122	0.199	-0.077	116.53
287	0.088	Not Measured	NA	Not Measured

Pearson's Correlation Between Differences (Column 4) and Averaged Rutgers Undergraduate Ratings (Column 5) is -0.39, P =0.0157 [Note IDs 63 and 287 with no differences between Brown's 1996 value and the correct 1996 value or that are missing the 1996 correct value are excluded from correlation analysis)

Table 4 – Differences Between Dr. Brown's RA_P and Correct (i.e. Self-Consistent with data set AD_P / M_P ratio) Summed FA in 2002 and Rutgers Undergrad Dance Evaluations Among 40 Selected Dancers

ID	Brown 2002 FA	Correct 2002 FA	Brown – Correct 2002 FA	Averaged Rutgers Undergraduate Student Ratings
15	0.103	0.211	-0.108	123.930
21	0.292	0.292	Same	98.150
23	0.135	0.159	-0.025	138.800
30	0.095	0.122	-0.028	100.580
33	0.301	0.301	Same	87.275
34	0.310	0.310	Same	89.450
38	0.067	0.118	-0.051	109.800
55	0.073	0.175	-0.101	129.575
63	0.322	0.322	Same	107.575
67	0.240	0.240	Same	100.725
68	0.136	0.144	-0.007	135.500
75	0.299	0.299	Same	73.930
86	0.137	0.146	-0.009	104.480
89	0.079	0.152	-0.073	118.980
94	0.333	0.233	0.100	105.800
103	0.347	0.347	Same	75.450
110	0.293	0.293	Same	112.980
113	0.343	0.303	0.040	109.480
115	0.335	0.435	-0.100	95.280
117	0.082	0.110	-0.027	110.750
119	0.265	0.265	Same	91.250
139	0.258	0.238	0.020	87.500
152	0.115	0.191	-0.076	121.400
162	0.075	0.173	-0.098	119.550
175	0.353	0.353	Same	67.880
182	0.132	0.149	-0.017	121.380
185	0.086	0.081	0.005	123.130
192	0.296	0.396	-0.100	99.380
194	0.124	0.152	-0.028	102.550
195	0.309	0.309	Same	113.775
197	0.093	0.139	-0.047	120.500
200	0.132	0.152	-0.020	115.500
203	0.115	0.164	-0.049	127.400
205	0.319	0.319	Same	108.350
206	0.264	0.264	Same	82.600
222	0.298	0.258	0.040	99.400
229	0.119	0.167	-0.048	121.130
235	0.315	0.315	Same	113.530
239	0.105	0.123	-0.018	116.530
287	0.097	0.169	-0.071	Not Measured

Pearson's Correlation Between Differences (Column 4) and Averaged Rutgers Undergraduate Ratings (Column 5) is -0.24, P =0.245 (Note – IDs 21, 33, 34, 63, 67, 75, 103, 110, 119, 175, 195, 205, 206 and

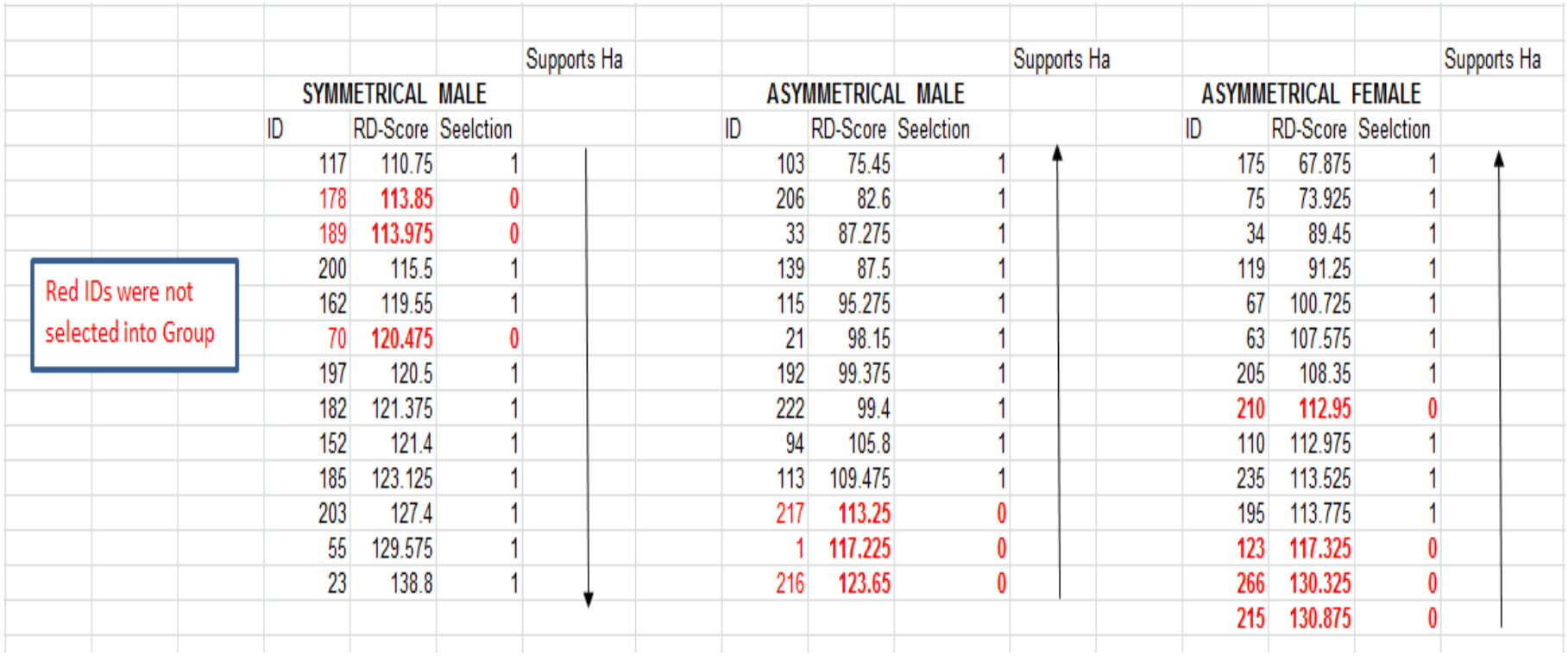
235 which have no differences between Browns 2002 value and correct 2002 value are excluded from correlation analysis)

Table 5. – Averaged Evaluator Dancer Ratings of 40 Jamaican Dancers in Dr. Brown's Dataset and Calculated by us and others from the file "individual_dance_ratings.xls" that Dr. Palestis received from Dr. Brown

ID	In Dr. Brown's Data Set	Calculations done by us and Others on raw scored in raw data initially sent to Dr. Palestis from Dr. Brown "individual_dance_ratings.xls"		
		By Us	By "Darine D"	By "Mean Dan"
15	48.90	63.43	63.43	62.20
21	30.39	29.32	29.32	27.80
23	62.21	70.83	70.83	69.02
30	40.28	42.91	42.91	40.97
33	41.04	27.12	27.12	26.07
34	27.93	28.02	28.02	27.47
38	51.11	29.40	29.81	28.46
55	59.78	78.44	78.44	76.92
63	37.45	37.03	37.03	36.07
67	39.36	38.58	38.58	37.33
68	50.91	73.35	73.35	72.41
75	30.45	29.42	29.83	28.48
86	37.50	26.04	26.03	25.02
89	47.65	47.16	47.16	45.05
94	54.00	60.58	60.58	58.64
103	31.38	29.48	29.88	29.30
110	52.74	65.36	65.36	64.52
113	43.47	62.22	62.22	61.01
115	35.66	25.92	25.92	24.58
117	52.37	45.60	45.60	43.84
119	31.58	27.98	28.10	27.20
139	40.99	52.44	52.68	52.68
152	63.75	63.35	63.53	63.12
162	55.98	34.89	35.26	34.12
175	17.03	15.78	15.78	14.96
182	70.09	70.58	70.70	70.25
185	30.65	30.95	30.95	29.75
192	50.53	58.46	58.46	56.97
194	32.05	14.59	14.50	13.66
195	44.67	52.28	52.52	51.18
197	59.21	68.79	68.49	68.49
200	63.79	63.34	68.49	62.70
203	55.29	55.64	55.82	55.10
205	37.05	30.22	30.32	28.96
206	23.73	24.11	24.29	23.19

222	41.06	39.88	40.21	39.43
229	40.33	57.10	57.31	56.57
235	37.58	54.65	54.88	54.18
239	40.88	41.70	42.03	40.40
287	65.73	71.37	71.49	70.57

Figure 1 – Diagram, of Rutgers Undergraduate Dance Ratings (RD) for Selected and non-Selected Candidate Dancers for symmetrical and asymmetrical males and asymmetrical females



NOTE – Arrow on Right Hand Side Points in Direction of RD Ratings that support Alternative Hypothesis