University at Buffalo
State University of New York

Department of Computer Science and Engineering

January 19, 2014

Judge Roberto Rivello
Nigel Freeman
Emil Sutovsky
FIDE-ACP Anti-Cheating Committee
FIDE Ethics Committee
Re: Complaint regarding FM Borislav Ivanov

Dear FIDE Committee Members:

I have been asked to provide a less-technical description of my involvement in testing allegations of cheating against Mr. Borislav Ivanov, which began on 2 January, 2013. I am an Associate Professor with tenure in Computer Science at the State University of NY at Buffalo, with a doctorate in Mathematics from Oxford University in 1986. Besides my long record of internationally-known research in computational complexity theory, I am co-manager of one of my field's prominent weblogs, and hence publicly immediately accountable. My research into statistical cheating claims began when I answered an open-channel request from Dr. Frederic Friedel of ChessBase on the day of Silvio Danailov's accusing letter in the 2006 World Championship match. Friedel asked for help with "how can we evaluate such accusations?", and that remains the larger subject here.

My statistical model, which I have presented in peer-reviewed papers and in talks available on my website, including the one to FIDE last October in Tallinn, is designed to test statistical hypotheses by generally-accepted procedures involving confidence intervals and so-called z-scores of deviation. The z-score is interpreted as odds against a suitably-framed "null hypothesis," which in mainstream fraud-detection applications means the sample comes (after all) from the normal run of business. Beyond the specific training data calibrating my model to the Elo scale, I have run vast numbers of games—including 50,000 played in 2013 alone—through computer analysis to determine what's *normal* in chess, so as to confirm the z-scores. The basic odds must be qualified by how the sample was selected. For instance, a z-score of 3.35 means about 2,500–1 odds against "normal," which sounds strong. But my 50,000 games include over 10,000 tournament performances (averaging 9–10 games each), so I should expect to find 4 z-scores so high just by what's normal. Hence if the only word in such a case is "the player matched the computer so much," it may mean nothing. But if there are independent words against a player, such as observations of behavior or breaking other rules, then this selection bias is avoided and the full force of 2,500–1 odds applies. Two years ago, at the time my work was written up in the New York Times and long before I heard anything about Mr. Ivanov, I wrote a webpage titled "The Parable of the Golfers" to explain this.

After seeing the Ivanov controversy on holiday on New Year's Eve, upon my return on 2 Jan. 2013, I ran my screening test that evening and my full test through the 4th. For the first time ever in my work, z-scores above 5.00 tumbled out of my program, to my amazement even topping 6.00 when I excluded the game when he was searched. After giving maximum benefit of doubt I reported 4.72 as my official result, which is just short of 1,000,000–1 odds. This is also what you get if you have **two** consecutive z-scores of 3.35 or higher,[1] which explains my private behavior (while silently watch-dogging events since 2009) of saying nothing unless I would see two such scores by the same player in close succession. The case of Sébastien Feller involves two scores near 3.00, of

[1]Using Stouffer's rule: $z = (z_1 + z_2)/\sqrt{2} = 6.70/1.414... = 4.74$.

Davis Hall Room 338, Box 602000, Buffalo, NY 14260-2000
Tel: (716) 645-3180 Fax: (716) 645-3464

course with supplementary evidence, and I had even written on the "Golfers" page that I did not expect to see above 3.50 from a single performance. The fact that Ivanov's 2012 Zadar Open hit the million-to-one target in one shot caused me to change this behavior. I felt the defense of chess required the risk of stepping forward, with my formal letter and report addressed to ACP and article "The Crown Game Affair" on said weblog. By so doing I was giving the benefit of scientific regularity and circumspection to supplement others' move-matching claims that had been made without due process.

In spring 2013 I was contacted by GM Kiril Georgiev after he lost a Rapid game to Ivanov. My test showed positive, but since it was only one game there were too few moves to put the z-score much over the minimum 2.00 conventionally needed to claim statistical evidence. None of his other games in Rapid events he won has come to light. However his six played games from the Veliko Tarnovo ("Old Capital") Open in May gave results near 3.00, and over 3.00 upon excluding games against players rated more than 200 Elo lower, per my allowed policy.

Now I am writing most specifically about his performances at the Blagoevgrad Grand Open last October and the Navalmoral Open last month. I tracked the former round-by-round, doing every game by every player in the screening tests. I noted to this committee just before my flight to Paris that after 3 rounds he was first in both move-matching (MM) and least average-error (AE). It is common for some player to lead both tests in the screening run since they are heavily correlated, but in a field of over 100, why should that player be the 14th-ranked Ivanov? As I reported technically a week ago, my full runs give mild positives (just above 2.00) in one or the other test. From Navalmoral only the last 3 of his 5 played games have been made available, while from Blagoevgrad there is anecdotal report of differing behavior in some games, including the two after he was forfeited against GM Maxim Dlugy. He is also distinguished at least by the behavior cited in his initial 4-month ban by the Bulgarian Chess Federation, so that the initial allowance for selection bias no longer applies. The extent to which observations made on his person at both events are admissible is a separate matter, but to that extent the statistical results have greater import.

In summary, these are the three particular considerations in this case that matter to the practice of using statistical evidence under consideration by FIDE:

1. The first tournament gave results that not only deviate significantly from normal, but deviate even compared to my whole record of deviations;
2. Subsequent tournaments, indeed all of them since May, show statistical positives;
3. There is now independent observational evidence from his last two events; this and statistical evidence strengthen each other.

I will be happy to answer any further questions and to provide technical results, besides those already communicated, upon request.

Yours sincerely

Dr. Kenneth W. Regan
(716) 645-4738, regan@buffalo.edu