Understanding the handicapping system

On the whole the new online handicapping system has been a success and there is a feeling that the operation of the system should be explained. This is an attempt to do that and I apologize if for some of you it is at a level which is too simplistic or for others too technical where I have introduced some very basic ideas in statistics and probability.

Preliminaries

It is important to realize that handicapping, in tennis or in any sport, is only an approximation (were the handicapping exact in horse-racing then all handicap races would end as a dead heat); there will always be variations on a player's performance due to internal factors such as food eaten the day before, level of concentration at work, bio-rhythms, etc. I would consider that if the handicap was within two points in tennis then we would be doing well.

The basis of the handicapping system in tennis is that a handicap difference of six points means that, were the game to be played at 15 – love, this would be an even game. If two players with a handicap difference of six points played level, then the expectation would be that the best player would win two out of three games.

For handicap differences below about ten, the odds given in the handicap tables are based on probability calculations using the six handicap difference as the unit of currency. For larger handicap differences the odds given are based on empirical results; i.e. on what seems to work in practice.

There will be some randomness in the match results. If player A has a better handicap than B then it does not mean that player A will always win against player B. However over a number of matches we expect that player A will win more often than not against player B.

This idea of randomness can be seen when we consider some coin-tossing experiments. If we toss a fair coin ten times then we would expect five heads and five tails. However this will not occur in practice; all the possibilities from all heads to all tails can be obtained, some outcomes being more likely than others. The full table of probabilities (rounded up, which is why the probabilities won't add up to 1 exactly) is:

No. of heads	0 or 10	1 or 9	2 or 8	3 or 7	4 or 6	5
Probability	0.001	0.01	0.04	0.12	0.21	0.25

For only a quarter of tosses we will obtain five heads and five tails; this will be the same probability as obtaining 3 of one and 7 of the other (twice 0.12). Twice out of a hundred we will obtain 9 of one and 1 of the other.

This idea of random outcome can be extended to simulating the outcomes of matches by coin tossing. Consider two players who are matched equally so that the probability of each winning a rest is 0.5; this is equivalent to tossing a fair coin to simulate the outcome of the rest. The probabilities for the various match outcomes are:

Outcome	Game to love	Game to 15	Game to 30	Deuce
Probability	0.125	0.250	0.3125	0.3125

This is quite surprising. A quarter of the time we will have a game to 15 between two equal players just due to randomness. Of course if a large number of games were played then we should obtain an equal proportion of rests, and games, won by each player. The short-term random effect is diluted by considering a large number of results.

Handicaps are not necessarily a transitive relationship. By this I mean that if player A is better than (i.e. wins more matches than loses against) player B and player B is better than player C then it is not necessarily true that player A is better than player C.

The basic model

The model that is used in practice is that the relative handicap difference between two players can be judged on the basis of the match result between the two players. The program then decides whether the relative handicap between the two players should remain the same or be adjusted to reflect the outcome; each player's handicap is moved by the same amount, one improving, the other worsening. It cannot decide whether one of the players is at the correct handicap and adjust the other player's handicap accordingly. If one player's handicap becomes incorrect because of this movement then other results for that player will bring the handicap back into line. The handicap is based on the outcome from many match results; a sort of consensus of all the matches played by the player; the more results the more accurate the handicap.

In calculating the handicap adjustment the computer will use the following data:

- The actual difference in handicaps between the two players; information which is currently on its database.
- The difference in handicaps at which the game was played (it may have been played level or the handicap used may have been rounded up to the nearest whole odds); this information is recorded with the match result.
- The match result.

The computer calculates the number of games won by each player and adds an extra game for winning a completed set; e.g. if the score was 6/4, 3/6, 3/5 then the first player would have won (6+1) + 3 + 3 = 13 games and the second player would have won 4 + (6+1) + 5 = 16 games; the first player has won 13 out of 29 games. Then the computer works out the effective handicap difference which is the difference between the handicap at which played and the actual difference in handicaps; e.g. if the actual handicap difference at the time the match was played was 4 and the match was played at 15-love (=+6) then the effective handicap difference is -2 (the negative sign indicating that the handicap at which played over-compensated the first player).

Because the first player was over-compensated the expectation is that he or she would win more of the games; in fact 16 out of the 29. The statistical conundrum is to decide how far from this ideal result one can deviate before the result becomes significant indicating that we should change the relative handicaps of the two players. In this example the first player would have registered a small loss with a consequent handicap adjustment of 0.2

Statistical interlude: *consider the table for the outcome of tossing a coin ten times for which the results are given in the table above; if we obtained ten heads from tossing a coin ten times, the probability of which is 0.001 or 1 in a thousand, could we infer that the coin is biased? It is so unlikely that we should be suspicious of the fairness of the coin. What could we infer if the outcome was 9 heads and 1tail; would we think that this is also unlikely to have occurred randomly and conclude that the coin is biased? Well such a result would only have occurred randomly with a probability of 0.01 and we could also conclude, but less convincingly, that the coin was biased. The conundrum is to decide at what point we should draw the line and accept some deviation from the average result. If we are too conservative in drawing the line then we will accept the result to be due to randomness whereas it may be due to bias; if we are too liberal then some actually random results may be taken to be due to bias.*

We, the HSRC, decided to have three outcomes; a big win or loss, a small win or loss, and a draw. A big win (or loss) would move the handicaps by 0.6 and will be applied whenever the outcome is less likely to occur than 1 in 10. A small win (or loss) would move the handicaps by 0.2 and will be applied whenever the outcome is less likely to occur than 1 in 3. Otherwise it is a draw. Most statisticians would consider these boundaries as very liberal; the justification for us to adopt such a policy is that these are not just rare results but the sum of many observations and also that it will facilitate handicap movement for those whose handicaps are not quite correct.

To return to the example above where 29 games are played the boundaries are:

Outcome	Big loss	Small loss	Draw	Small win	Big win
Games won out of 29	0 to 12	13 to 14	15 to 18	19 to 20	21 to 29

With 13 games won out of 29, player A would suffer a small loss and the handicap would be worse by 0.2; the opponent's handicap would improve by 0.2. It may seem to be a small shift from a big loss at 12 games to a draw at 15 out of 29 games played but the probability of such events occurring randomly are very different.

There is a mistaken belief that if a player only played one opponent who was worse then the handicaps would increasingly differ. This is not the case since as the handicaps would increase in difference so the better player would have to win by an increased margin to record a win or big win; eventually the two handicaps will level out and stabilize.

There is also a mistaken belief that a match win should count as a handicap win. Were this the case then there would be no handicap movement at all unless the worse player won. For handicap purposes the quality of the win is important; if the object of the match is to win then any win will suffice whereas if the object is to improve or retain one's handicap then the win must be a convincing one.

Extensions to the basic model

There are a number of extensions to the basic model:

1. Home court advantage. This is a parameter that is supposed to make allowance for the advantage of one player being familiar with the court. Currently the value is 3 handicap points (based on past results from the Brodie and Field Trophy matches) but once we have more data, then it may be changed.
2. Limit on the effective handicap. Once a certain limit, currently 6, is reached for the effective difference in handicap then the calculations become too sensitive and results are declared null unless the worse player wins in which case this is counted as a big win (and big loss for the better player). It is important that the handicap at which the game is played bridges the gap between the two players' handicaps.
3. Volatility. Those players whose handicaps are not a true reflection of their actual ability (e.g. beginners, returning from an enforced absence due to injury or being sent to work to an uncivilized location bereft of tennis courts, intensive coaching, *anno domini*) are deemed to be volatile (this does not reflect on their temperament on court). The results of a match between a volatile playing and a non-volatile player are then used to estimate the handicap of the volatile player using the non-volatile player as a yardstick.
4. For tournament matches the handicap movement is increased, e.g. 1.2 for a big win in a three-set match.

Conclusion

This is a model and is the best that we have been able to produce so far. However we are continually seeking to improve our methods and will be publishing more details on the handicapping web-site. Please write in with any helpful comments or queries., The more results that are input the better the mapping of handicaps to player's abilities. Remember that the whole point of handicapping is to provide enjoyable and competitive games between players. There is no point in having an inflated handicap as a status symbol nor in winning tournaments because one's handicap is too high.

John Trapp
Chair of Handicapping and Rankings Sub-Committee
5 July 2003