

# Comment gagner un tournoi d'échecs ?

Jean Hausser

17 mars 2005

## Résumé

Ce rapport fait appel à une analyse multivariée pour tenter de déterminer une stratégie optimale à adopter au cours d'un tournoi d'échecs. Pour cela, il s'appuie sur des données d'un tournoi récent comportant des joueurs de niveau international.

## Introduction

Si on en exclut l'aspect ludique, on peut considérer que le but d'un tournoi d'échecs est de déterminer quel participant est le meilleur joueur. En tournoi, la plupart des parties d'échecs durent 3 ou 4 heures, mais cela peut aller jusqu'à 6 heures. Etant donné le nombre de participants — d'une vingtaine à plusieurs centaines — il n'est pas envisageable que tout le monde rencontre tout le monde. Il s'agit donc de déterminer le meilleur joueur en organisant un nombre minimum de parties. En pratique, le nombre de parties dépend du nombre de participants attendus et de l'importance de la compétition.

Ainsi, un tournoi est une suite de 7 ou 9 rondes. Une ronde est un ensemble de parties qui ont lieu simultanément, en parallèle. Une partie est un affrontement entre deux joueurs choisis par un processus appelé appariement. Dans les tournois de faible importance, chaque participant joue 7 rondes, ou 9 dans les grandes compétitions. Chaque partie gagnée rapporte un point au vainqueur et 0 au perdant. En cas de match nul, les deux joueurs reçoivent  $1/2$  point. A la fin du tournoi, on additionne les points de chacun et on classe les participants d'après le nombre de points obtenus. Les ex-aequo sont ensuite départagés par un système de départage fixé à l'avance par l'organisation du tournoi. Celui-ci est basé sur des critères comme le rang moyen des joueurs rencontrés pendant le tournoi, le rang de la première partie perdue sur les 9 rondes, le cumul des points à la fin de chaque ronde, etc... A chaque ronde, l'adversaire rencontré n'est pas choisi au hasard. L'appariement — qui rencontre qui ? — des joueurs est déterminé par le classement provisoire et obéit à des règles spécifiques, constantes au cours du tournoi et fixées par l'organisation. Quelque soit le système choisi, le rang des joueurs rencontrés dépend directement du classement provisoire et donc de la performance du joueur. De plus, on ne peut pas rencontrer le même joueur deux fois au cours du tournoi.

Mon expérience personnelle de joueur d'échecs en compétition m'amène à penser que la dernière ronde ne change généralement plus beaucoup le classement et que les premières sont les plus déterminantes pour le classement. Cependant, perdre une partie n'est pas éliminatoire. Certains joueurs perdent même volontairement leur première partie par forfait pour se ménager un tournoi plus facile et « faire la course dans le peloton » dans le but de détrôner le joueur de tête dans les dernières parties. D'autres encore quittent le tournoi si ils perdent leur première partie. Pour un joueur, il est donc crucial de savoir quelles sont les parties qu'il est le plus important de gagner. Pour tenter de répondre à cette question, on considère les résultats du 2<sup>e</sup> Open International de Balagne de février 2005 qui a rassemblé 111 participants. Les résultats sont disponibles sur le site de la Fédération Française des Echecs.

## Traitement des données

**Première méthode** On fait un copier/coller du tableau dans un fichier texte. Puis on le traite à l'aide de la commande suivante, qui fait appel aux outils de traitement de texte standards d'Unix.

```
sed 's/\t/,/g' < tournoi.txt | sed 's/ //g' | \  
awk 'BEGIN {FS=","} {print $1,$4,$8,$9,$10,$11,$12,$13,$14,$15,$16}' | \  
tr '<' '-' | tr '>' '+' | sed 's/[BN]//g' | grep -v 'EXE' | \  
sed 's/[1-9][0-9]*/1/g' | sed 's/-[1-9][0-9]*/0/g' | \  
sed 's/[1-9][0-9]*/0.5/g' | sed 's/ / 0 /g' | sed 's/ / 0 /g' \  
> tourn-ok1.txt
```

On a ici éliminé les joueurs qui ont été exempts (sans adversaire) au cours du tournoi. En sortie, on obtient un fichier texte de 107 individus et 11 variables prêt à importer dans R. Le premier champs contient le classement final du joueur. Le second est son classement ELO (un entier positif indiquant le niveau du joueur au niveau international). Les 9 champs restants contiennent le résultat de chaque partie : 0 en cas de perte, 1 en cas de gain, 0.5 en cas de match nul.

```
> tourn=read.table("~/work/Chessell/tourn-ok1.txt", h=F)  
> hist(apply(tourn[,3:11], 1, sum), freq=F, col="lightblue", main="",  
      xlab="Score", ylab="")  
> lines(density(apply(tourn[,3:11], 1, sum)))  
> shapiro.test(apply(tourn[,3:11], 1, sum))
```

Shapiro-Wilk normality test

```
data:  apply(tourn[, 3:11], 1, sum)  
W = 0.9323, p-value = 3.773e-05
```

```
> ks.test(2*apply(tourn[,3:11], 1, sum),"pbinom", 9*2, 0.5)
```

One-sample Kolmogorov-Smirnov test

```
data:  2 * apply(tourn[, 3:11], 1, sum)  
D = 0.1989, p-value = 0.0004206  
alternative hypothesis: two.sided
```

```
> mean(apply(tourn[,3:11], 1, sum))  
[1] 4.313084  
> abline(v=mean(apply(tourn[,3:11], 1, sum)))
```

L'histogramme des scores finaux de la figure 1 page 3 montre que la répartition est centrée autour de 4.5 points, c'est à dire la moitié des 9 possibles. Ce résultat est attendu : à chaque ronde,  $\lceil \frac{107}{2} \rceil = 54$  parties sont jouées. A chaque ronde, 54 points sont donc répartis entre les 107 joueurs du tournoi, ce qui fait environ 0.5 points par joueur et par partie, ou encore  $0.5 * 9$  rondes = 4.5 sur le tournoi. En pratique, la moyenne des scores (ici 4.3) est légèrement inférieur à 4.5 car certains joueurs arrivent en cours de tournoi ou le quittent avant la fin ce qui fausse le calcul. Dans tous les cas, la répartition n'est ni binomiale, ni normale.

On effectue une analyse en composantes principales sur ces données après centrage et réduction.

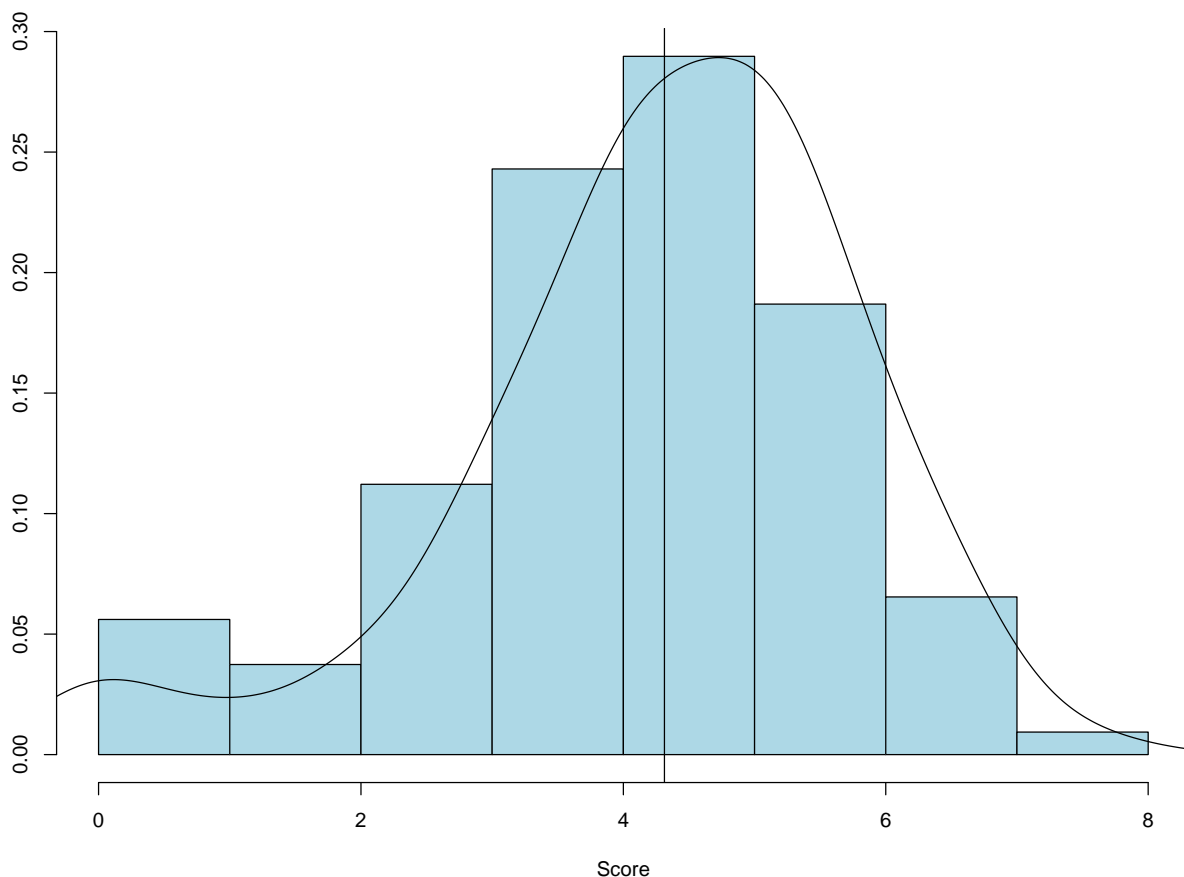


FIG. 1 – Histogramme des scores à l'issu du tournoi

```

> library("ade4")
> attr(tourn, "names")=c("Classement", "ELO", "R1", "R2", "R3", "R4", "R5",
  "R6", "R7", "R8", "R9")
> pca1=dudi.pca(tourn[,3:11])
Select the number of axes: 2
> s.value(pca1$li, tourn$Classement, method="greylevel")
Warning message:
X11 used font size 8 when 9 was requested
> s.arrow(2*pca1$co, add.plot=T)

```

Comme on peut le voir sur la figure 2 page 4, il n'apparaît pas de structuration claire sur ce plan factoriel. Les meilleurs joueurs sont confondus avec les autres et les projetés des variables ne permettent pas de les regrouper de façon cohérente. On a alors tenté de répartir les 15 premiers joueurs, les 15 derniers et les autres en trois classes, mais aucune organisation claire n'a pu être mise en évidence par ce moyen.

**Deuxième méthode** Si la première approche a échoué, cela doit tenir au prétraitement des données, et plus précisément au choix de ne conserver que les résultats de chaque partie (0, 1 ou 0.5). L'inconvénient de n'avoir conservé que ces données est qu'on perd beaucoup d'information concernant la force du joueur rencontré. Pourtant, une partie gagnée contre un joueur fort, bien

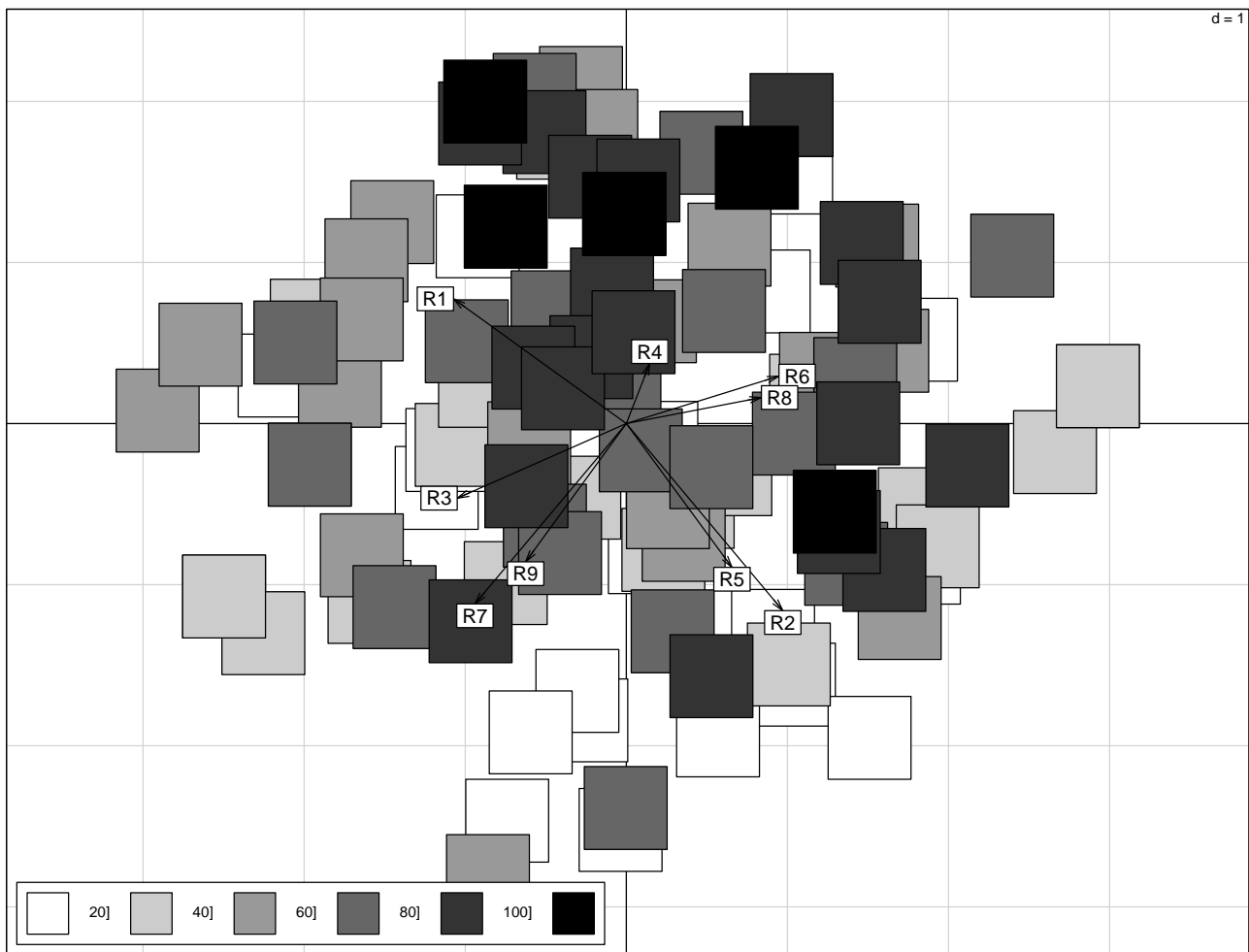


FIG. 2 – Premier plan factoriel. Chaque carré correspond à un joueur.

classé, « vaut » plus qu’une victoire contre le dernier du tournoi alors que l’information « gain, match-nul ou perte » confond les deux cas.

L’idéal serait de transformer les données en donnant une note algébrique à chaque partie. Par exemple, une victoire contre un joueur fort rapporterait plus qu’une victoire contre un joueur faible. Cela vaudrait également pour une défaite : mieux vaudrait perdre contre un joueur fort que contre un joueur faible. Mais, fait-on mieux la démonstration de sa valeur en faisant match-nul contre un joueur fort, ou en gagnant contre un joueur faible ? Comme il est difficile de trancher cette question, on élimine l’information « gain, match-nul ou perte » et on ne conserve que le rang des joueurs rencontrés à chaque ronde :

```
sed 's/\t/,/g' < tournoi.txt | sed 's/ //g' | \
awk 'BEGIN {FS=","} {print $1,$4,$8,$9,$10,$11,$12,$13,$14,$15,$16}' | \
tr '<' '-' | tr '>' '+' | sed 's/[BN]//g' | grep -v 'EXE' | \
sed 's/[=+-]//g' | sed 's/ / 112 /g' | sed 's/ / 112 /g' > tourn-ok2.txt
```

Cette commande extrait du tableau brut le classement final du joueur, son classement ELO, et le classement final du joueur rencontré à chaque ronde. On a comme précédemment éliminé les individus qui ont été exemptés au cours du tournoi c’est à dire sans adversaire, en raison du nombre impair de joueurs. Restent donc 107 joueurs. Comme les absents ne sont pas appariés, une absence à une partie est ici arbitrairement assimilée à une partie jouée contre le joueur 112,

qui est un joueur fictif mais qui indique une mauvaise performance puisqu'il est « dernier des derniers » par son classement.

On refait l'ACP sous R :

```
> tourn=read.table("~/work/Chessel/tourn-ok2.txt", h=F)
> attr(tourn, "names")=c("Classement", "ELO", "R1", "R2", "R3", "R4", "R5",
  "R6", "R7", "R8", "R9")
> pca1=dudi.pca(tourn[,3:11])
Select the number of axes: 2
> s.value(pca1$li, tourn$Classement, method="greylevel")
> s.class(pca1$li, as.factor(c(rep("A", 20), rep(NA, 107-20))),
  col="red", add.plot=T)
> s.class(pca1$li, as.factor(c(rep(NA, 20), rep("B", 107-20-20), rep(NA, 20))),
  col="blue", add.plot=T)
> s.class(pca1$li, as.factor(c(rep(NA, 20), rep(NA, 107-20-20), rep("C", 20))),
  col="green", add.plot=T)
> s.arrow(2*pca1$co, add.plot=T)
```

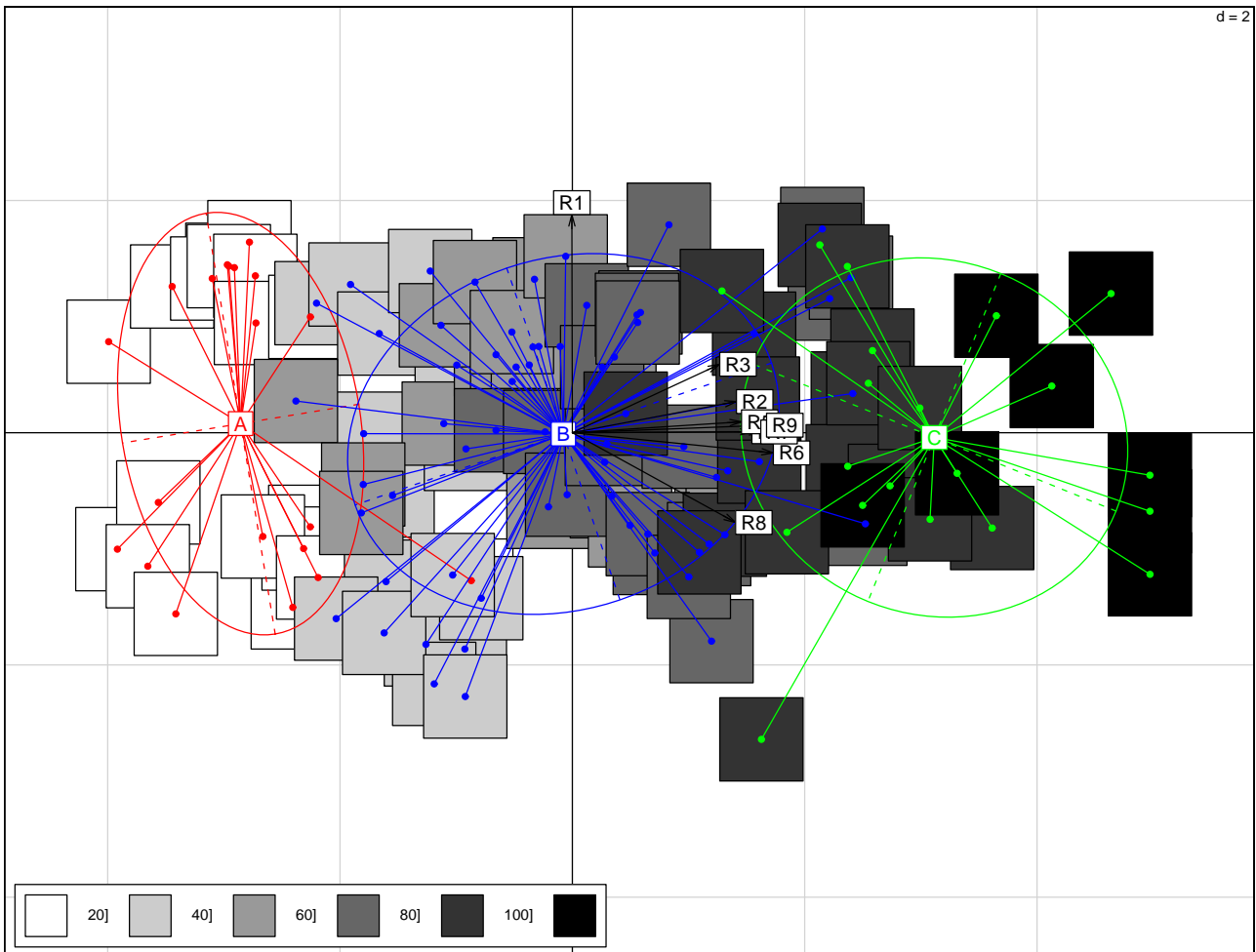


FIG. 3 – Deuxième plan factoriel. Chaque carré correspond à un joueur.

Cette fois, une organisation apparaît clairement (voir figure 3 page 5). Les vingt meilleurs joueurs sont regroupés dans la classe A, les vingt derniers en C et les autres en B. Ces classes se répartissent précisément *sur* la première composante principale, dans le sens opposé puisqu'un

classement petit indique une bonne performance. C'est donc la première composante principale qui détermine largement le classement final.

## Conclusion

Revenons à la question posée dans l'introduction. Quelles rondes sont les plus importantes dans un tournoi? La réponse apparaît clairement dans le plan factoriel et sur le cercle des corrélations associé.

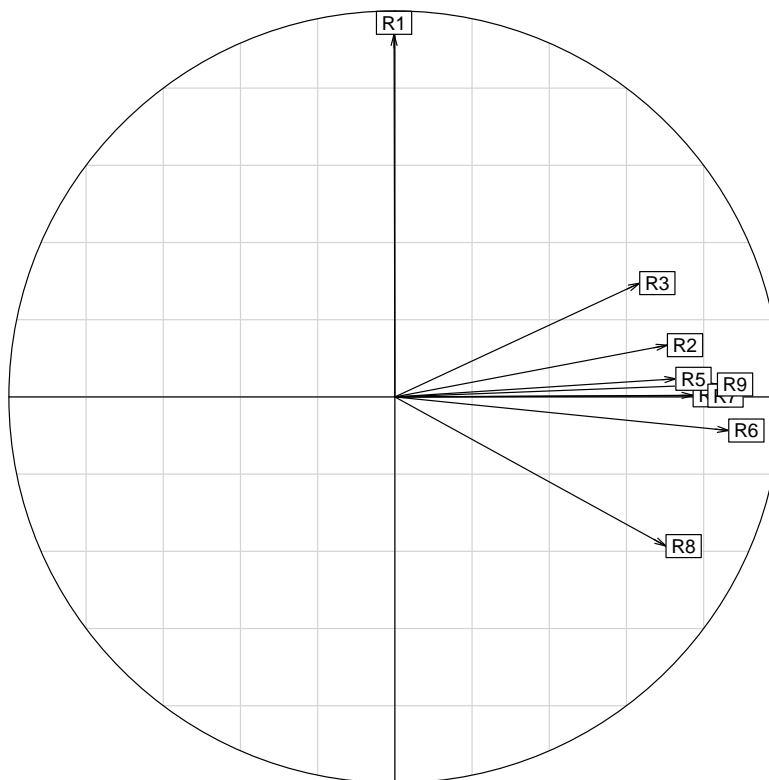


FIG. 4 – Cercle de corrélation des variables

Sur la figure 4 page 6, on voit que R1 est orthogonale à la première composante principale qui indique la performance du joueur dans le tournoi. La variable R1 ne détermine donc aucunement le classement final des joueurs. La stratégie consistant à perdre la première partie pour se ménager un tournoi plus facile et arriver reposé en fin de tournoi pour « renverser » le joueur de tête est donc aussi payante que celle qui consiste à faire le tournoi en tête. On voit aussi que les dernières rondes comme R9 et R7 sont très corrélées à la première composante principale. Les dernières rondes ont donc un rôle déterminant pour le classement final du tournoi. En outre, et à l'exception de la R1, toutes les variables sont corrélées positivement à la première composante. Hormis la R1, toutes les rondes influent donc sur le classement final. Enfin, les résultats sont reproductibles sur d'autres tournois de type et d'importance comparables.

Au final, je n'ai jamais été un grand adepte de cette stratégie, principalement parce que ma fierté m'interdisait de me laisser vaincre par un joueur plus faible, ou de perdre contre un joueur fort sans me battre. Mais au vu des résultats de cette étude, je compte bien changer de stratégie. Pourquoi se fatiguer quand on peut arriver au même résultat avec moins d'efforts?