Soccer players with darker skin are more likely to get a red card

Authors: J. Ullrich[1]*, E. Schlüter[2], C. Spörlein[3], A. Glenz[1]

Affiliations

[1]University of Zurich, Department of Psychology.

[2]University of Giessen, Institute of Sociology.

[3]University of Bamberg.

*Correspondence to: j.ullrich@psychologie.uzh.ch.

Abstract

Racism is a known problem among soccer supporters, but it is unclear to what extent even professional referees are biased against soccer players with darker skin. We examined the relative likelihood of players with darker skin-tone to receive a red card using life-history data of players from four major European leagues. Based on generalized linear mixed models we found that a player with the darkest skin-tone was 1.38 times as likely to receive a red card as a player with the brightest skin-tone. The result could be due to a referee bias, a greater propensity of players with darker skin-tone to commit severe fouls, or a combination of both. We also examined country means of implicit and explicit racial bias as a moderator of this effect at the level of referees' country of origin, but the results were inconclusive.

One Sentence Summary

Soccer players with darker skin are more likely to get a red card.

Results

Data preparation

The two ratings of skin-tone were averaged and rescaled to the range from 0 (corresponding to the original value of 1) to 1 (corresponding to the original value of 5). This was done so that the difference between brightest and darkest players could later on be read off the regression coefficients directly. The new variable was called avgrate01.

Cases were excluded if they had missing values on skin-tone-rating, meanIAT or meanExp (listwise deletion) because we wanted to perform all analyses (including research question 2) on the same set of cases.

Consistent with the original research question 1 (Are soccer referees more likely to give red cards to dark skin toned players than light skin toned players?) we were interested in the likelihood of a player receiving a red card in a single game. The original response variable redCards is uninterpretable because the number of games a player has seen a given referee varies. Therefore we disaggregated the data (one game per row, redCards appear as 1s in the first n=redCards rows per player). This was possible because it does not matter in which of, for example, 3 games a player who received 1 red card in 3 games received the red card. It is sufficient that this player has three observations (three rows) associated with him, one of them indicating a red card. We used the binomial error distribution because – after disaggregation – our response variable specifies the occurence of an event in a single game, coded 0 and 1.

Modeling strategy

Using the statistical software R (Version 3.1.1; R Core Team, 2014), we estimated generalized linear mixed models (function `glmer` in R package lme4, Version 1.1-7; Bates, 2010; Bates, Maechler, Bolker, & Walker, 2014). With this technique we can estimate the desired effects while accounting for random variance of the effects across players, referees, and referees' country of origin. The crowdstorming data are different from standard multilevel data (e.g., where employees are members of only one team), inasmuch as they are not nested but cross-classified – player A can have multiple games with the referee A, but player B can have multiple games with the same referee A. The package lme4 can deal with such a data structure.

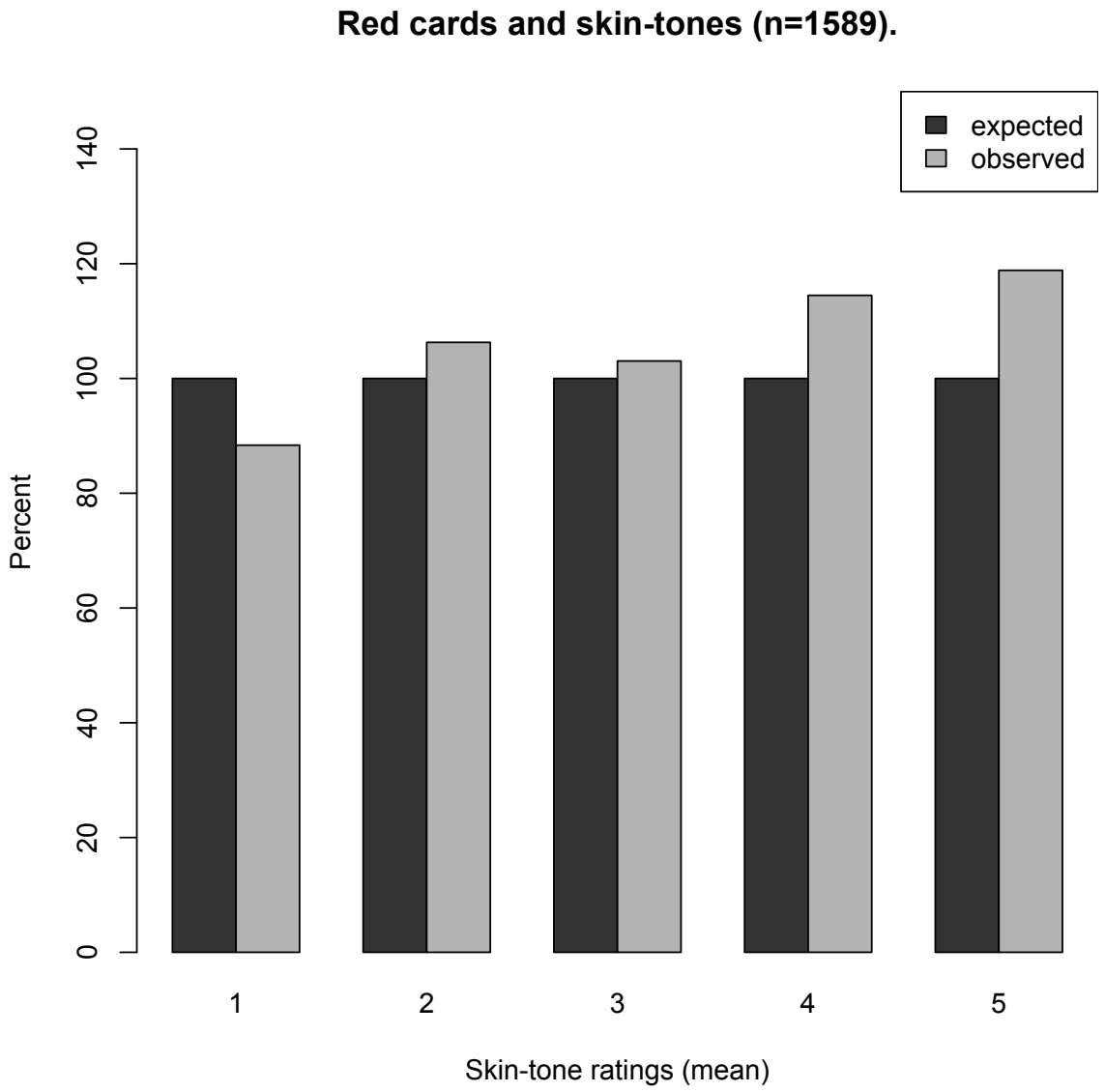
In all models we predicted the event of getting a red card in a single game (0 = no, 1 = yes) based on the skin-tone variable avgrate01, exploring the usefulness of different random effects and moderator variables. We did not use any covariates such as player position and decided to stick with this approach even though reviewers of our approach suggested that we should do so. As already noted in the project description by Silberzahn, Martin, Uhlmann, & Nosek, the data cannot be used for causal inference. Thus, if the goal is to come up with a generalizable descriptive statement (i.e., effect size), it does not matter why a player ends up getting more red cards (e.g., being a tall, heavy defense player). In fact, such information when included as covariates might even bias the result.

Research Question 1: Are soccer referees more likely to give red cards to dark skin toned players than light skin toned players?

To answer research question 1, we first performed a simple visual comparison (see Figure 1 below) of the expected and observed relative frequencies of red cards across the five categories of skin-tone (rounding off the values on avgrate01 to avoid small cell frequencies). Note that the

observed relative frequencies were calculated from the subset of player-game-combinations in which a player received a red card (n = 1589).

Figure 1.



As can be seen in Figure 1, there seems to be a linear increase of the percentage of redCards as we go from the lowest category to the highest category in skin-tone. In order to test this relationship more formally, we compared four models:

```
gm0 <- glmer(redCards ~ 1+(1 |playerShort) + (1|refNum),
             family = binomial, data = data.games.nona, nAGQ=0)
```

Model gm0 is the reference model that includes only the intercept (a fixed intercept and two random intercepts). The random term for refNum corrects for individual differences in referees' propensity to draw a red card. The random term for playerShort captures the individual differences in players' propensity to receive a red card.

```
gm1 <- glmer(redCards ~ 1+avgrate01+(1 |playerShort) +
 (1|refNum),family = binomial, data = data.games.nona,nAGQ=0)
```

Model gm1 adds a fixed effect of skin-tone (avgrate01, transformed so that 0 represents the brightest skin-tone and 1 the darkest).

```
gm2 <- glmer(redCards ~ 1+avgrate01+(1 |playerShort) +
 (1+avgrate01|refNum),family = binomial, data =
 data.games.nona,nAGQ=0)
```

Model gm2 adds a random effect of skin-tone across referees.

```
gm3 <- glmer(redCards ~ 1+avgrate01+(1 |playerShort) +
 (1|refNum) + (1+avgrate01|refCountry),
              family = binomial, data = data.games.nona,nAGQ=0)
```

Model gm3 adds a random effect of skin-tone across referees' countries of origin. Table 1 presents the results of these model fits.

Comparing the model fits shown in the upper part of Table 1, model gm1 was better than model gm0 according to AIC and a likelihood ratio test, suggesting that skin-tone predicts the odds of getting a red card. Model gm2 was not superior to model gm1 as indicated by larger values on AIC and BIC, suggesting that the effect of skin-tone does not vary across referees (variance value close to zero). However, model gm3 was better than than model gm1 according to AIC and a likelihood ratio test, indicating some variance of the effect of skin-tone across referees' countries of origin. Thus, we interpret the coefficients from gm3.

In the column for model gm3, the fixed intercept gives the log-odds of getting a red card in a game for a typical white player (brightest skin-tone = 0). Converting this to a probability value through the expression `exp(-5.8334)/(1+exp(-5.8334))`, we obtain .0036, that is, a white player has a chance of 0.36% to get a red card in a single game. This is close to the simple arithmetic mean calculated on the variable redCards only for players in the category brightest skin-tone (.0039). The fixed effect for skin-tone (avgrate01) indicates the increase in the log-odds as we go from the category of players with the brightest skin to the category of players with the darkest skin, assuming a linear increase. Converting the sum of the intercept and avgrate01 to a probability value, we obtain .0050, that is, a player with the darkest skin-tone has a chance of 0.5% to get a red card in a single game. This is close to the simple arithmetic mean for players with the darkest skin-tone (.0051).

The odds of getting a red card were 1.38 times higher for a player from the darkest skin-tone category compared with a player from the brightest skin-tone category. The lower bound of the 95% confidence interval (Wald method) was `exp(.0917)= 1.10` and the upper bound

was `exp(.5582)= 1.75`. For comparison, the log-odds of .3641 corresponds to a small effect in the effect size d metric `(.325*sqrt(3)/pi) = .18`.

Research Question 2: Are soccer referees from countries high in skin-tone prejudice more likely to award red cards to dark skin toned players?

At the level of referees' countries of origin, we visualized the relationship between the indicator of implicit racial bias (i.e., meanIAT) and the random effects of avgrate01 from the model gm3 in a scatterplott with a loess regression line:

Figure 2.

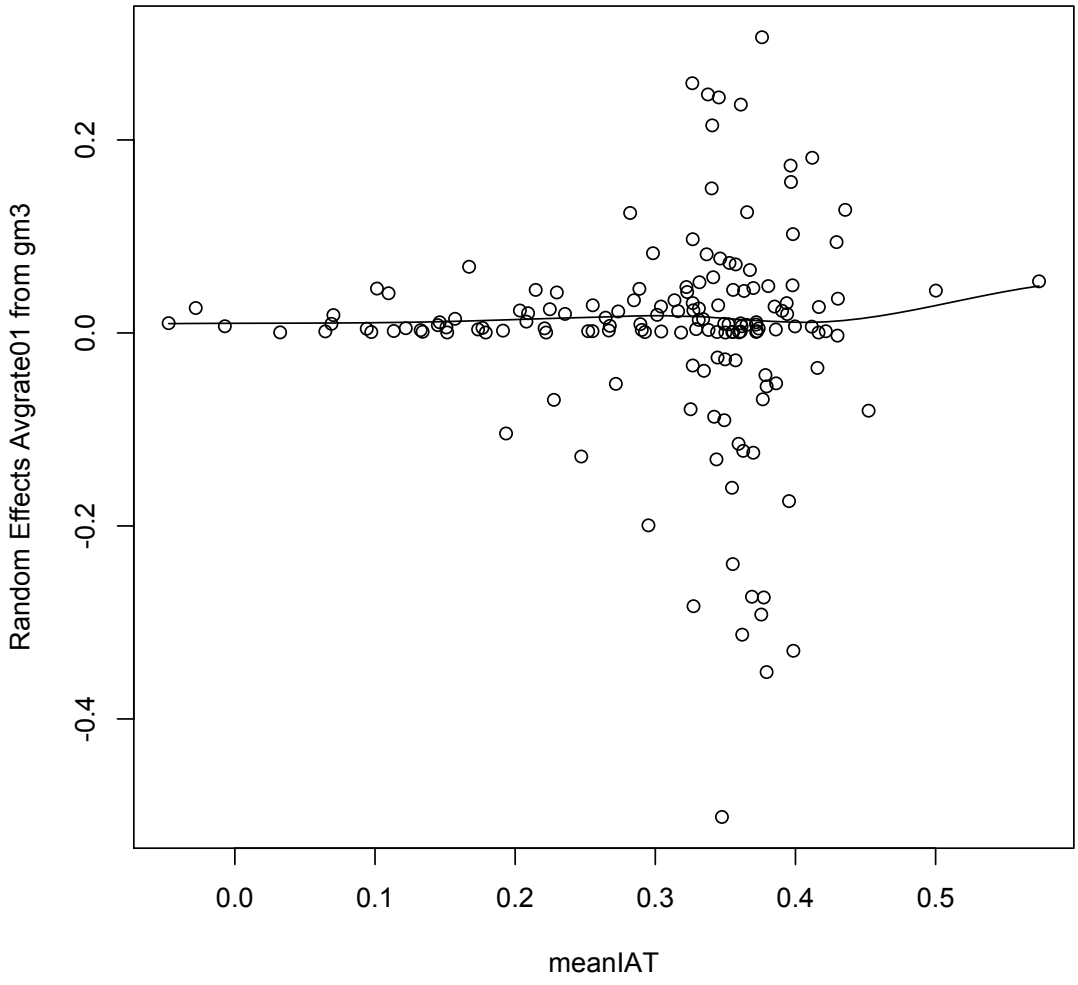
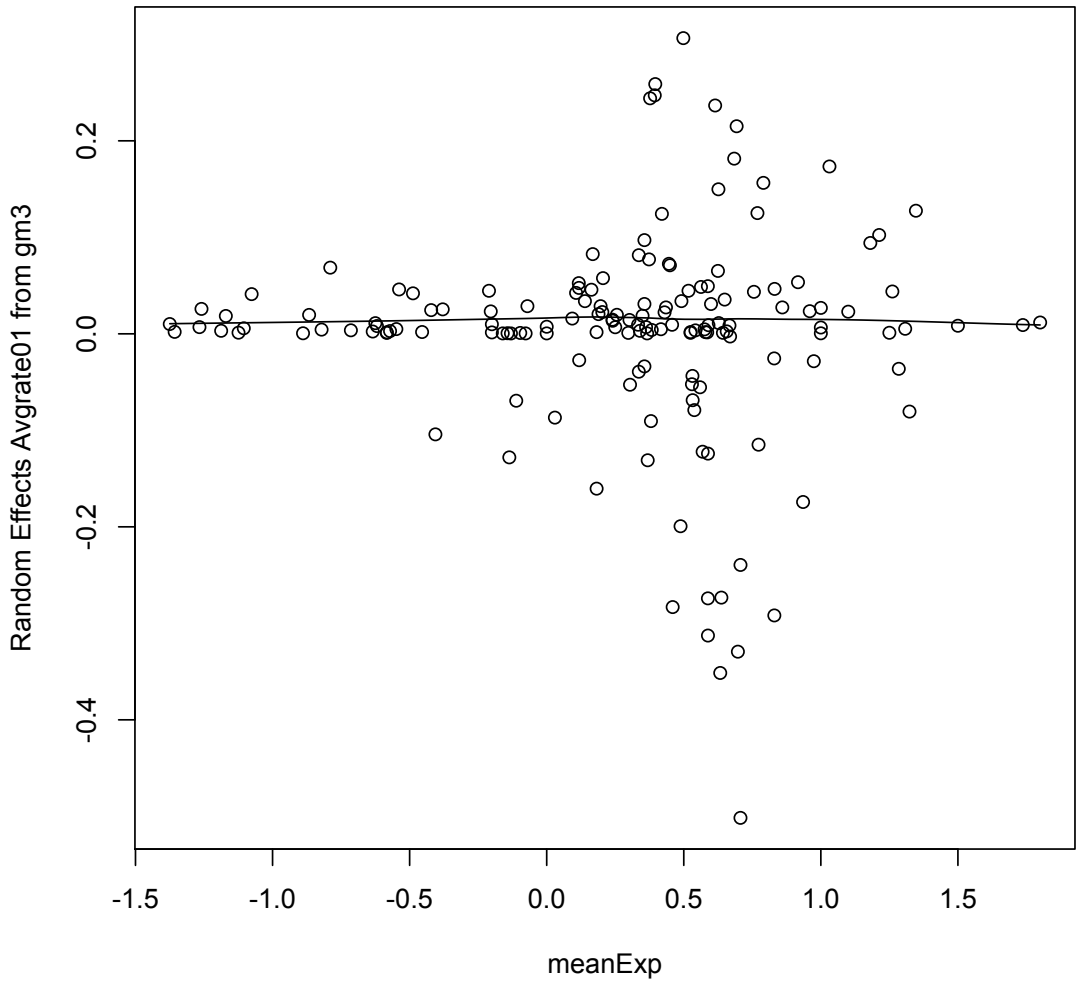


Figure 2 reveals no evidence of a relationship between the IAT country means for referees' country of origin and the effect of skin-tone, linear or otherwise. The effect of the interaction between avgrate01 and meanIAT was .55 (*SE* = 3.80, *p* = .89). The odds-ratio was

1.73, that is the odds-ratio darkest skin over brightest skin (research question 1) is multiplied by 1.73 for a unit-increase in meanIAT. The lower limit of the 95% confidence interval (Wald method) for the interaction effect between skin-tone and meanIAT was `exp(-6.8974) = 0.001` and the upper limit as `exp(8.0006)= 2983`. In other words, there is considerable uncertainty as to the true effect size.

Next, we visualized the relationship between the indicator of explicit racial bias (i.e., meanExp) and the random effects of avgrate01 from gm3:

Figure 3.



Consistent with the results for implicit skin-tone bias, Figure 3 does not reveal any evidence of a relationship between explicit skin-tone bias and the random effects of avgrate01 from gm3. The effect of the interaction between avgrate01 and meanExp was .33 (*SE* = .55, *p* =

.55). The odds-ratio was 1.39, that is the odds-ratio darkest skin over brightest skin (research question 1) is multiplied by 1.39 for a unit-increase in meanIAT.

The lower limit of the 95% confidence interval (Wald method) for the interaction effect between skin-tone and meanExp was `exp(-.7537) = .47` and the upper limit was `exp(1.4119)= 4.10`. In other words, there is considerable uncertainty as to the true effect size.

Conclusion

Our analyses suggest that a darker skin-tone increases the likelihood that a player will get a red card. Compared with players from the brightest skin-tone category, players from the darkest skin-tone category are 1.38 times as likely to get a red card. This effect is consistent with several explanations. Referees might be biased against players with darker skin-tones. Alternatively, players with darker skin-tones might be more prone to commiting severe fouls that warrant a red card. Both of these factors may in fact be responsible for this effect. To sort out these possibilities it would be useful to conduct an experiment with professional referees and show identical fouls committed by players with darker or brighter skin-tone. As for the lack of evidence on the role of racial bias at the level of referees' country of origin, we suspect that margin of error was high because most referees and players come from countries with similar levels of racial bias scores (especially for the IAT) so that the lower and upper end of the bias scale was not well represented.

Tables

Table 1.

Model Fit	gm0	gm1	gm2	gm3
AIC	20379	20371	20374	20333
BIC	20411	20414	20439	20409
logLikelihood	-10186	-10182	-10181	-10160
deviance	20373	20363	20361	20319
Variances of Random Effects				
Intercept|playerShort	.3249	.3215	.3214	.2796
Intercept|refNum	.1308	.1298	.1616	.0689
avgrate01|refNum	–	–	.0210	-
Intercept|refCountry	–	–	–	.3950
avgrate01|refCountry	–	–	–	.0677
Fixed Effects				
Intercept	-5.4991	-5.5931	-5.5969	-5.6280
avgrate01 (Standard Error)	–	.3229 (.10)**	.3295 (.10)**	.3250 (.12)**

Note: avgrate01 = averaged ratings of player skin-tone (0=brightest, 1=darkest); playerShort = players; refNum = referees; refCountry = referees' country of origin; ** p < .01

Data and Output

```r
# crowdstorming analyses ullrich, schlüter, spörlein, glenz
# R script

require(lme4)

data<-read.csv("crowdstorming.csv")


# player-referee combo variable
data$p_ref <- paste(data$playerShort,data$refNum,sep=".")

# average skintone rating
data$avgrate <- apply(cbind(data$rater1,data$rater2),1,FUN=function(x)
 mean(x,na.rm=TRUE))

# repeat rows for each game
data.games <-
 as.data.frame(matrix(ncol=ncol(data),nrow=sum(data$games)))
names(data.games) <- names(data)
for (col in 1:ncol(data))
{
 data.games[,col] <- rep(data[,col],data$games)
}

# build vector of redCards per game
redCards.games <- c()
for (i in 1:nrow(data))
{
 x <- c(rep(1,data$redCards[i]),rep(0,data$games[i]-data$redCards[i]))
    redCards.games <- c(redCards.games,x)
}
data.games$redCards <- redCards.games

# exclude cases with missing values
data.games.nona<-subset(data.games,!is.na(avgrate) &
   !is.na(meanIAT))

# rescale skin-tone-rating to range 0,1
data.games.nona$avgrate01 <- (data.games.nona$avgrate-1)/4

# use redcard-games only for plot of expected and observed frequencies
data.red <- data.games.nona[data.games.nona$redCards > 0,]
```

```
# save frequencies
x <- table(as.integer(data.games.nona$avgrate))
y <- table(as.integer(data.red$avgrate))

# Chi-Square test: Compare observed frequencies with expected
 probability
null.probs <- x / sum(x)
chi <- chisq.test(y,p=null.probs)
print(chi)
print(data.frame(ratings=as.numeric(names(x)),expected=round(as.vector
 (chi$expected),1),observed=as.vector(chi$observed)))

cols <- c("grey20","grey70")
b <- barplot(t(cbind(chi$expected / chi$expected * 100,chi$observed /
 chi$expected * 100)),
             beside=TRUE,
             col=cols,
             xlab="Skin-tone ratings (mean)",
             ylab="Percent",
             ylim=c(0,150),
             main=paste("Red cards and skin-tones
 (n=",sum(chi$observed),").",sep=""))
     legend("topright",legend=c("expected","observed"),fill=cols)

# glmer fits for research question 1 (gm3 best model). Use option
 nAGQ=0 for faster processing by the cost of a slightly reduced
 accuracy:

gm0 <- glmer(redCards ~ 1+(1 |playerShort) + (1|refNum),
              family = binomial, data = data.games.nona,nAGQ=0)

gm1 <- glmer(redCards ~ 1+avgrate01+(1 |playerShort) + (1|refNum),
              family = binomial, data = data.games.nona,nAGQ=0)

gm2 <- glmer(redCards ~ 1+avgrate01+(1 |playerShort) +
 (1+avgrate01|refNum),
              family = binomial, data = data.games.nona,nAGQ=0)

gm3 <- glmer(redCards ~ 1+avgrate01+(1 |playerShort) + (1|refNum) +
 (1+avgrate01|refCountry),
              family = binomial, data = data.games.nona,nAGQ=0)

> summary(gm3)
Generalized linear mixed model fit by maximum likelihood (Adaptive
```

```
  Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
 Family: binomial  ( logit )
Formula: redCards ~ 1 + avgrate01 + (1 | playerShort) + (1 | refNum) +
    (1 + avgrate01 | refCountry)
   Data: data.games.nona

     AIC      BIC   logLik deviance df.resid
 20333.0  20408.8 -10159.5  20319.0   372873

Scaled residuals:
    Min      1Q  Median      3Q     Max
-0.2185 -0.0706 -0.0614 -0.0544 28.0611

Random effects:
 Groups      Name        Variance Std.Dev. Corr
 refNum      (Intercept) 0.06886  0.2624
 playerShort (Intercept) 0.27963  0.5288
 refCountry  (Intercept) 0.39504  0.6285
             avgrate01   0.06777  0.2603   -0.93
Number of obs: 372880, groups:  refNum, 2967; playerShort, 1585;
 refCountry, 155

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.6280     0.1146  -49.11  < 2e-16 ***
avgrate01     0.3250     0.1190    2.73  0.00632 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
          (Intr)
avgrate01 -0.600

> confint(gm3,parm=2,method="Wald")
               2.5 %    97.5 %
avgrate01 0.09173171 0.5581725

# Research Question 2a:
# scatterplot of meanIAT and random effects of avgrate01 at the level
 of refCountry:

refag<-
 aggregate(data.games.nona$meanIAT,list(data.games.nona$refCountry),mea
 n)
```

```
colnames(refag)<-c("refCountry","meanIAT")

refag$ranefavgrate01 <- NA
refag$ranefavgrate01 <- ranef(gm3,drop=FALSE)$refCountry[,2]

scatter.smooth(refag$meanIAT,refag$ranefavgrate01,xlab="meanIAT",ylab=
 "Random Effects Avgrate01 from gm3")


gm4 <- glmer(redCards ~ 1+avgrate01*meanIAT+(1 |playerShort) +
 (1|refNum) + (1+avgrate01|refCountry),
              family = binomial, data = data.games.nona,nAGQ=0)

> summary(gm4)
Generalized linear mixed model fit by maximum likelihood (Adaptive
  Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
 Family: binomial  ( logit )
Formula: redCards ~ 1 + avgrate01 * meanIAT + (1 | playerShort) + (1 |
    refNum) + (1 + avgrate01 | refCountry)
   Data: data.games.nona

     AIC      BIC   logLik deviance df.resid
 20335.7  20433.1 -10158.8  20317.7   372871

Scaled residuals:
    Min      1Q  Median      3Q     Max
-0.2182 -0.0706 -0.0614 -0.0543 28.0945

Random effects:
 Groups      Name        Variance Std.Dev. Corr
 refNum      (Intercept) 0.06891  0.2625
 playerShort (Intercept) 0.27955  0.5287
 refCountry  (Intercept) 0.39332  0.6272
             avgrate01   0.06415  0.2533   -0.94
Number of obs: 372880, groups:  refNum, 2967; playerShort, 1585;
 refCountry, 155

Fixed effects:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        -6.1994     0.9900  -6.262 3.79e-10 ***
avgrate01           0.1427     1.3204   0.108    0.914
meanIAT             1.6137     2.7839   0.580    0.562
avgrate01:meanIAT   0.5516     3.8006   0.145    0.885
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) avgr01 menIAT
avgrate01   -0.752
meanIAT     -0.993  0.754
avgrt01:IAT  0.744 -0.996 -0.753

> confint(gm4,parm=4,method="Wald")
                      2.5 %   97.5 %
avgrate01:meanIAT -6.89735 8.000634

# Research Question 2b:
# scatterplot of meanExp and random effects of avgrate01 at the level
 of refCountry:

refag<-
 aggregate(data.games.nona$meanExp,list(data.games.nona$refCountry),mea
 n)
colnames(refag)<-c("refCountry","meanExp")

refag$ranefavgrate01 <- NA
refag$ranefavgrate01<-ranef(gm3,drop=FALSE)$refCountry[,2]

scatter.smooth(refag$meanExp,refag$ranefavgrate01,xlab="meanExp",ylab=
 "Random Effects Avgrate01 from gm3")

gm5 <- glmer(redCards ~ 1+avgrate01*meanExp+(1 |playerShort) +
 (1|refNum) + (1+avgrate01|refCountry),
              family = binomial, data = data.games.nona,nAGQ=0)

> summary(gm5)
Generalized linear mixed model fit by maximum likelihood (Adaptive
  Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
 Family: binomial  ( logit )
Formula: redCards ~ 1 + avgrate01 * meanExp + (1 | playerShort) + (1 |
    refNum) + (1 + avgrate01 | refCountry)
   Data: data.games.nona

     AIC      BIC   logLik deviance df.resid
 20334.9  20432.4 -10158.5  20316.9   372871

Scaled residuals:
    Min      1Q  Median      3Q     Max
```

```
-0.2182 -0.0707 -0.0614 -0.0543 28.1403

Random effects:
 Groups      Name        Variance Std.Dev. Corr
 refNum      (Intercept) 0.06876  0.2622
 playerShort (Intercept) 0.27929  0.5285
 refCountry  (Intercept) 0.39990  0.6324
             avgrate01   0.07321  0.2706   -0.93
Number of obs: 372880, groups:  refNum, 2967; playerShort, 1585;
 refCountry, 155

Fixed effects:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        -5.7399     0.2473 -23.210   <2e-16 ***
avgrate01           0.1984     0.2785   0.713    0.476
meanExp             0.1822     0.4054   0.450    0.653
avgrate01:meanExp   0.3291     0.5524   0.596    0.551
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) avgr01 menExp
avgrate01   -0.693
meanExp     -0.883  0.682
avgrt01:mnE  0.592 -0.900 -0.713

> confint(gm5,parm=4,method="Wald")
                         2.5 %   97.5 %
avgrate01:meanExp -0.7536732 1.411863
```

References and Notes

Bates, D.M. (2010). lme4: Mixed-effects modeling with R. Available from http://lme4.r-forge.r-project.org/lMMwR/lrgprt.pdf

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7, <URL: http://CRAN.R-project.org/package=lme4>.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.