

### Hautton und rote Karten im Fußball

Förster, Martin

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

**Empfohlene Zitierung / Suggested Citation:**

Förster, M. (2016). *Hautton und rote Karten im Fußball*. (Schriftenreihe für erweiterte Replikationen, Crowdsourcing und empirische Theorieüberprüfung, 1/2016). Flensburg: Europa-Universität Flensburg, Interdisziplinäres Institut für Umwelt-, Sozial- und Humanwissenschaften. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-458130>

**Nutzungsbedingungen:**

Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>

**Terms of use:**

This document is made available under a CC BY-NC-ND Licence (Attribution-Non Commercial-NoDerivatives). For more information see:

<https://creativecommons.org/licenses/by-nc-nd/4.0>

## **Schriftenreihe für erweiterte Replikationen, Crowdsourcing und empirische Theorieüberprüfung**

---

Nr. 1 (2016)

Martin Förster

**Hutton und rote Karten im Fußball**

### **Kontakt / Impressum**

Martin Förster  
Europa-Universität Flensburg  
Interdisziplinäres Institut für Umwelt-, Sozial- und Humanwissenschaften  
Auf dem Campus 1  
24943 Flensburg

Email: [ifv@foerster-m.de](mailto:ifv@foerster-m.de)

# Hautton und rote Karten im Fußball

Martin Förster <sup>a,\*</sup>

<sup>a</sup>Europa-Universität Flensburg, Interdisziplinäres Institut für Umwelt-, Sozial- und Humanwissenschaften, Auf dem Campus 1, 24943 Flensburg

---

## Zusammenfassung

Im Fußball neigen Schiedsrichter dazu, Spielern mit dunklem Hautton eher eine rote Karte zu geben als Spielern mit hellem Hautton.

---

In einem sogenannten 'Crowdsourcing Analytics'-Projekt (Silberzahn et al., 2015) ging es den Autoren darum, den gleichen Datenbestand hinsichtlich der selben Forschungsfragen von mehreren Teams analysieren zu lassen. Das Forschungsinteresse hierbei war, wie unterschiedliches methodisches Vorgehen die Ergebnisse beeinflusst<sup>1</sup>. 29 Teams sollten den Fragen nachgehen, ob, erstens, Fußballschiedsrichter dazu neigen, Spielern mit dunklem Hautton eher eine rote Karte zu geben als Spielern mit hellerem Hautton, und zweitens, ob dieser Effekt ggf. von Vorurteilen im Herkunftsland der Schiedsrichter moderiert wird. Die Teams waren frei in der Wahl ihrer Methoden zur Untersuchung der Fragestellungen und frei in der Wahl der Kovariaten, alle erhielten den selben Datenset. Letztlich ging jedes Team anders vor. 20 Teams konnten einen signifikanten positiven Effekt ermitteln, die anderen neun Teams konnten keine signifikanten Effekte finden.

Dieser Beitrag strebt keine Replikation von einem der 29 Teamreports an, sondern eine weitere 'Lösungsmöglichkeit' mit den bereitgestellten Daten.

## Daten

Die Fälle des Datensets sind Spieler-Schiedsrichter-Dyaden, d.h. die Kombination bzw. das Aufeinandertreffen von Spieler und Schiedsrichter in Spielen. Es gibt keine 'leeren' Dyaden. Wenn also ein Spieler und ein Schiedsrichter nie in einem Spiel aufeinandergetroffen sind, dann gibt es für diese Kombination

---

\*Email: [ifv@foerster-m.de](mailto:ifv@foerster-m.de)

<sup>1</sup>Projektbeschreibung, Daten, Ergebnisse, und Diskussionen sind abrufbar unter: <https://osf.io/gvm2z/>

auch keine Dyade. Die Identifikationsvariablen jeder Dyade sind eine Identifikation des Spielers (`playerShort`, `player`) und eine Identifikation des Schiedsrichters (`refNum`). Zudem ist die Anzahl der Spiele in einer Dyade (`games`) erfasst. Weiterhin sind Variablen vorhanden zu Größe (`height`), Gewicht (`weight`), und (Feld-) Position (`position`) des Spielers, zur Anzahl der Siege (`victories`), unentschiedenen Spiele (`ties`) und Niederlagen (`defeats`) der Spieler in der jeweiligen Dyade, zur Anzahl der geschossenen Tore der Spieler in der Dyade (`goals`), Anzahl gelber (`yellowCards`), gelb-roter (`yellowReds`), und roter Karten (`redCards`) des Spielers in der Dyade, ein Verweis auf ein Fotobild des Spielers (`photoID`), Ratings zum Hautton der Spieler von zwei Ratern (`rater1`, `rater2`), eine Identifikation des Herkunftslandes des Schiedsrichters als numerischer (`refCountry`) und alpha-numerischer (`Alpha_3`) Code, und Variablen zur Vorurteilsmessung bezüglich des Hauttons bezogen auf das Herkunftsland des Schiedsrichters. Es gibt ebenfalls Variablen zum Verein, in welchem die Spieler zum Zeitpunkt der Datenerhebung spielten (`club`), zum Land, in dessen Liga der Verein spielt (`leagueCountry`) - ebenfalls zum Zeitpunkt der Datenerhebung, und zum Geburtsdatum der Spieler (`birthday`). Das Datenset umfasst alle Fußballspieler, die während der Saison 2012-2013 in der höchsten Liga in England, Deutschland, Frankreich, oder Spanien spielten. Es ist zu beachten, dass lediglich das Sampling der Spieler nach diesen Kriterien (Saison und Liga) erfolgte, die Informationen der Dyaden aber die gesamte Laufbahn der entsprechenden Spieler umfasst: "Thus the data entails a period of multiple years from a player's first professional match until the date this data was acquired (June 2014)" (Silberzahn et al., 2015, p.6).

Der Datenset enthält  $N_{dyads} = 146028$  Fälle (Dyaden), welche die Kombination von  $N_{players} = 2053$  Spielern und  $N_{referees} = 3147$  Schiedsrichtern umfasst.

## Neigen Fußballschiedsrichter dazu, Spielern mit dunklem Hautton eher eine rote Karte zu geben als Spielern mit hellerem Hautton?

Diese Forschungsfrage zielt auf das Handeln der Schiedsrichter: Geben Schiedsrichter dunkelhäutigen Spielern eher eine rote Karte als hellhäutigen Spielern? So, wie diese Frage formuliert ist, kann sie nicht ohne weiteres mit den verfügbaren Daten beantwortet werden. Der Grund dafür liegt im Sampling. Wir reflektieren die Hypothese: die unabhängige Variable ist Hautton der Spieler, die abhängige Variable ist das Ereignis 'rote Karte'. Die Fälle sind Spieler-Schiedsrichter-Dyaden. Für eine vereinfachte Darstellung des Problems nehmen wir an, dass sich der Hautton ebenfalls als dichotomes Ereignis erfassen lässt ('hellhäutig' und 'dunkelhäutig'). Ein stochastischer Zusammenhang dieser Ereignisse ließe sich dann mit solchen Maßen ermitteln, die auf einer Vierfeldertafel basieren. Mit dieser Vierfeldertafel (Tabelle 1) könnte der Zusammenhang ggf. auch hinsichtlich seiner Richtung beschrieben werden (also ob dunkelhäutige Spieler mehr oder weniger rote Karten bekommen).

Das Datenproblem bezüglich dieser Vierfeldertafel besteht darin, dass die

Tabelle 1: Vierfeldertafel Hautton und rote Karten

Hautton (x)	rote Karte (y)		$\Sigma$
	ja	nein	
hell	I	II	A
dunkel	III	IV	B

Randsummen (A und B) unbekannt sind. Somit kann die Vierfeldertafel keine Aussage über einen Zusammenhang machen, weil eine belastbare Aussage dazu sich auf die Zeilenprozentage beziehen müsste.

Es mag nicht unmittelbar einsichtig sein, weshalb wir behaupten, dass die Randsummen unbekannt sind. Um das klar zu machen, müssen wir uns fragen, was denn die Fälle sind, die in dieser Vierfeldertafel erfasst werden müssen. Dafür können wir uns vorstellen, dass der Zusammenhang für einen bestimmten Schiedsrichter untersucht werden soll. Die Frage wäre dann also: Gibt es einen Zusammenhang zwischen dem Hautton der Spieler und der Vergabe roter Karten durch den Schiedsrichter? Die Fälle müssten hier alle Spieler sein, die eine rote Karte von diesem Schiedsrichter hätten bekommen können. Dies nun aber bedeutet, dass das Sampling auf Spielebene hätte durchgeführt werden müssen, sodass alle Spieler der jeweiligen Spiele erfasst worden wären. In den Daten gibt es jedoch keine Informationen zur Identifikation von Spielen.

Aber selbst wenn diese Informationen in den Daten vorhanden wären, muss bedacht werden, dass rote Karten in einem Spiel eher seltene Ereignisse sind. Weiterhin kann nicht ausgeschlossen werden, dass sich die Wahrscheinlichkeit des Ereignisses 'rote Karte' mit jedem bereits vorher im entsprechenden Spiel eingetretenen Ereignis ändert.

Allerdings sehen wir eine Möglichkeit, diese Probleme zu umgehen, indem wir die Forschungsfrage leicht abändern. Voraussetzung dafür ist, dass es darum geht, einen allgemeinen Zusammenhang zwischen x und y zu ermitteln - also nicht nur auf spezielle Schiedsrichter beschränkt. Dann wiederum müsste sich der Zusammenhang zu dem Phänomen aggregieren, dass dunkelhäutige Spieler relativ zu ihren Spielen mehr rote Karten erhalten als hellhäutige Spieler. Mit der Struktur der vorliegenden Daten lässt sich dieser Zusammenhang überprüfen.

Um den Unterschied bei der Forschungsfrage noch einmal deutlich zu formulieren: für die ursprüngliche Forschungsfrage ist  $y :=$  'Schiedsrichter vergibt rote Karte'; die abgewandelte Forschungsfrage definiert  $y :=$  'Spieler erhält rote Karte'. Die Randsummen für die abgewandelte Forschungsfrage sind bekannt, weil hier die Situation, in welcher der Schiedsrichter eine rote Karte vergibt - also das Spiel und die Zusammensetzung der Spielerschaft bezüglich A und B - irrelevant ist.

Tabelle 2: Häufigkeiten (Dyaden) Rating Hautton

Rating Hautton (x)	f	p
(1) $0 \leq x < 0.5$	95725	0.7681
(2) $0.5 \leq x \leq 1$	28896	0.2319
N	124621	



Tabelle 3: Deskriptive Statistik

	N	$\hat{\mu}$	$\hat{\sigma}$	Min	Max
Rating Hautton (x)	124820	0.2836945	0.2880686	0	1
rote Karten (y)	146028	0.0082808	0.0102498	0	0.2
Vorurteilsmessung (Exp)	145865	0.4520255	0.2174689	-1.375	1.8

## Operationalisierung

Die primäre unabhängige Variable, Hautton der Spieler, wurde durch zwei Rater codiert. Höhere Werte bedeuten eine dunklere Hautfarbe des Spielers. Wir haben die zwei Ratingwerte gemittelt. Das Wertespektrum reicht von 0 (beide Rater schätzen den Hautton als sehr hell ein) bis 1 (beide Rater schätzen den Hautton als sehr dunkel ein). Insgesamt überwiegen Ratings von hellen Hauttönen (Tabelle 2, Tabelle 3).

Insofern wir uns mit der Reformulierung der Forschungsfrage dazu entschlossen haben, nicht Spieler-Schiedsrichter-Dyaden als Fälle zu verwenden, sondern die Spieler, mussten wir die abhängige Variable - die Anzahl roter Karten, die der jeweilige Spieler bekommen hat - etwas umstrukturieren. Zunächst wurden für jede Dyade die roten und gelb-roten Karten addiert. Diese Summe wurde dann ins Verhältnis zu der Anzahl der Spiele gesetzt, die in der Dyade gezählt wurden. Diese relative Häufigkeit roter Karten wurde dann je Spieler über die Dyaden gemittelt<sup>2</sup>.

## Analyse

Für einen ersten Blick verwenden wir den Rho-Korrelationskoeffizienten. Der ermittelte Zusammenhang ( $\rho = 0.0796$ ,  $p < 0.0001$ ) ist positiv und signifikant. D.h. dunkelhäutige Spieler bekommen mehr rote Karten als hellhäutige Spieler.

<sup>2</sup>Eine alternative Operationalisierung der abhängigen Variable wäre gewesen, die Anzahl der roten Karten eines Spielers (über alle entsprechenden Dyaden) ins Verhältnis zu setzen zu der Anzahl der Spiele eines Spielers (über alle entsprechenden Dyaden) - also nicht die Anteile der roten Karten über die Dyaden, in denen der Spieler beteiligt ist, zu mitteln. Damit wäre jedoch die Annahme nötig, dass die Anteile hellhäutiger und dunkelhäutiger Spieler, denen ein Schiedsrichter in den Spielen begegnet, in etwa gleich groß sind. Wenn diese Annahme nicht erfüllt ist, bestünde bei dieser alternativen Operationalisierung die Gefahr, dass ein sogenanntes Simpson-Paradox auftritt (Blyth, 1972).

Tabelle 4: Schätzungen Modell 1

$y_i = \hat{b}X_i + \hat{a} + \hat{u}_{referee} + e_i$	$\hat{b}100$	SE100	T	$p$
Rating Hautton	0.2055	0.0088	22.9856	0.0000
Spiele (Dyade)	-0.0030	0.0008	-5.9226	0.0000
Gewicht (Spieler)	0.0102	0.0004	29.1033	0.0000
Anteil Siege	-0.8092	0.0583	-13.7780	0.0000
Anteil Niederlagen	0.2805	0.0710	3.8820	0.0000
Tore (Spieler)	-0.0879	0.0211	-4.1182	0.0000
Anteil gelbe Karten (Spieler)	5.8881	0.0371	158.7621	0.0000
a	-0.4409	0.0551	-8.1040	0.0000
B	0.2564			0.0000
$N_{dyads} = 124067$		$N_{referees} = 2980$		
	$\hat{\sigma}^2100$	SE100	Wald Z	$p$
e	0.0078	0.0000	247.9223	0.0000
u	0.0000	0.0000	4.9616	0.0000

Diese bivariate Analyse liefert uns nun aber lediglich den Befund eines totalen Effekts ohne Kontrolle auf weitere Variablen. Insofern wir die Forschungsfrage allerdings etwas strenger verstehen, fragen wir nach dem direkten und kontrollierten Effekt des Hauttons auf 'rote Karte', welchen wir zunächst mit Modell 1 (Tabelle 4) schätzen.

An dieser Stelle können wir feststellen, dass die bis hier ermittelten Koeffizienten des Hauttons auf signifikanten positiven Zusammenhang mit 'rote Karte' verweisen. D.h. wir können die erste Forschungsfrage mit 'ja' beantworten: dunkelhäutige Spieler erhalten mehr rote Karten als hellhäutige Spieler.

## Neigen Fußballschiedsrichter aus Ländern, in denen die Vorurteile bezüglich des Hauttons groß sind, dazu, Spielern mit dunklem Hautton eher eine rote Karte zu geben?

### Operationalisierung

Für die Messung von Vorurteilen (durchschnittlich in dem Land, aus welchem der Schiedsrichter einer Dyade kommt) hinsichtlich des Hauttons der Spieler wird ein 'Explicit Bias Score' (racial thermometer task, Exp) verwendet (Tabelle 3). Der ebenfalls mit den Daten gelieferte 'Implicit Bias Score' (IAT) hat eine sehr viel geringere Streuung (Variationskoeffizienten:  $V_{IAT} = 2.3140 < V_{Exp} = 7.6925$ ) und scheint mithin ungeeignet als Prediktor, insofern die potentiell mögliche Erklärungsleistung eben durch die eigene Varianz gewissermaßen

Tabelle 5: Schätzungen Modell 2

$y_i = \hat{b}X_i + \hat{a} + \hat{u}_{referee} + e_i$	$\hat{b}100$	SE100	T	$p$
Rating Hautton	0.2351	0.0176	13.1968	0.0000
Exp	0.0817	0.0186	4.4487	0.0000
Rating×Exp	-0.0658	0.0345	-1.8406	0.0635
Spiele (Dyade)	-0.0028	0.0008	-5.6019	0.0000
Gewicht (Spieler)	0.0102	0.0004	29.1446	0.0000
Anteil Siege	-0.8171	0.0584	-13.8773	0.0000
Anteil Niederlagen	0.2820	0.0711	3.9118	0.0000
Tore (Spieler)	-0.0895	0.0211	-4.2270	0.0000
Anteil gelbe Karten (Spieler)	5.8767	0.0372	157.9034	0.0000
a	-0.4756	0.0558	-8.6659	0.0000
B	0.2562			0.0000
$N_{dyads} = 123914$		$N_{referees} = 2969$		
	$\hat{\sigma}^2100$	SE100	Wald Z	$p$
e	0.0078	0.0000	247.7694	0.0000
u	0.0000	0.0000	4.9857	0.0000

beschränkt wird. Allerdings sehen wir auch davon ab, IAT dennoch zur Kontrolle in das Modell zu integrieren, weil die Korrelation von IAT mit Exp doch beachtlich ist ( $r = 0.7786$ ,  $p < 0.0001$ ) und wir damit (Multikollinearität) verzerrte Schätzer riskieren würden.

## Analyse

Modell 2 (Tabelle 5) kann keinen signifikanten Interaktionseffekt des Vorurteilsindikators mit dem Hautton ermitteln. D.h. die Vorurteile in einem Land bezüglich des Hauttons moderieren nicht den Effekt des Hauttons.

## Befunde und Diskussion

Es kann bestätigt werden, dass Spieler mit dunklem Hautton eher bzw. mehr rote Karten bekommen als Spieler mit hellem Hautton. Der ermittelte Effekt ist direkt und wird nicht durch Vorurteile, die in dem Herkunftsland des Schiedsrichters bestehen, moderiert. Die Stärke dieses Effekts in Odds-Ratios ausgedrückt beträgt  $OR = 1.2835$ , mit einem 95%-Konfidenzintervall  $CI_{95} = [1.2491, 1.3189]$ . Der Einfluss des Hauttons besteht neben und unabhängig von den Einflüssen, die ausgehen von: der Erfahrung der Spieler (Anzahl Spiele), der individuellen physischen Fähigkeit zum 'groben' Spiel (Gewicht), der emotionalen Belastung der Spieler (Anteil Siege, Niederlagen, und Tore), und dem tatsächlichen Hang zum 'groben' Spiel der jeweiligen Spieler (Anteil gelber Karten).

## Literatur

Blyth, Colin R. (1972): On Simpson's Paradox and the Sure-Thing Principle. *Journal of the American Statistical Association*, 67(338):364–366.

Silberzahn, Raphael; Eric Luis Uhlmann; Dan Martin; Pasquale Anselmi; Frederik Aust; Eli C. Awtrey; Stepan Bahník; Feng Bai; Colin Bannard; Evelina Bonnier; Rickard Carlsson; Felix Cheung; Garret Christensen; Russ Clay; Maureen A. Craig; Anna Dalla Rosa; Lammertjan Dam; Mathew H. Evans; Ismael Flores Cervantes; Nathan Fong; Monica Gamez-Djokic; Andreas Glenz; Shauna Gordon-McKeon; Tim Heaton; Karin Hederö Eriksson; Moritz Heene; Alicia Hofelich Mohr; Kent Hui; Magnus Johannesson; Jonathan Kaldimos; Erikson Kaszubowski; Deanna Kennedy; Ryan Lei; Thomas Andrew Lindsay; Silvia Liverani; Christopher Madan; Daniel Molden; Eric Molleman; Richard D. Morey; Laetitia Mulder; Bernard A. Nijstad; Bryson Pope; Nolan Pope; Jason M. Prenoveau; Floor Rink; Egidio Robusto; Hadiya Roderique; Anna Sandberg; Elmar Schlueter; Felix S; Martin Sherman; S. Amy Sommer; Kristin Lee Sotak; Seth Spain; Christoph Sporlein; Tom Stafford; Luca Stefanutti; Susanne Täuber; Johannes Ullrich; Michelangelo Vianello; Eric-Jan Wagenmakers; Maciej Witkowiak; SangSuk Yoon und Brian A. Nosek (2015): Many analysts, one dataset: Making transparent how variations in analytical choices affect results.