Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Fair Coins Tend to Land on the Same Side They Started: Evidence from 350,757 Flips

František Bartoš, Alexandra Sarafoglou, Henrik R. Godmann, Amir Sahrani, David Klein Leunk, Pierre Y. Gui, David Voss, Kaleem Ullah, Malte Zoubek, Franziska Nippold, Frederik Aust, Felipe Fontana Vieira, Chris-Gabriel Islam, Anton J. Zoubek, Sara Shabani, Jonas Petter, Ingeborg B. Roos, Adam Finnemann, Aaron B. Lob, Madlen F. Hoffstadt, Jason Nak, Jill de Ron, Koen Derks, Karoline Huth, Sjoerd Terpstra, Thomas Bastelica, Magda Matetovici, Vincent L. Ott, Andreea S. Zetea, Katharina Karnbach, Michelle C. Donzallaz, Arne John, Roy M. Moore, Franziska Assion, Riet van Bork, Theresa E. Leidinger, Xiaochang Zhao, Adrian Karami Motaghi, Ting Pan, Hannah Armstrong, Tianqi Peng, Mara Bialas, Joyce Y.-C. Pang, Bohan Fu, Shujun Yang, Xiaoyi Lin, Dana Sleiffer, Miklos Bognar, Balazs Aczel & Eric-Jan Wagenmakers

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2025, VOL. 00, NO. 0, 1–10: Applications and Case Studies
https://doi.org/10.1080/01621459.2025.2516210

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Fair Coins Tend to Land on the Same Side They Started: Evidence from 350,757 Flips

František Bartoš[a], Alexandra Sarafoglou[a], Henrik R. Godmann[a], Amir Sahrani[a], David Klein Leunk[a], Pierre Y. Gui[b], David Voss[b], Kaleem Ullah[b], Malte Zoubek[c,*], Franziska Nippold[b], Frederik Aust[a], Felipe Fontana Vieira[d,**], Chris-Gabriel Islam[e,f], Anton J. Zoubek[g], Sara Shabani[h], Jonas Petter[a], Ingeborg B. Roos[i], Adam Finnemann[a,j], Aaron B. Lob[b,k], Madlen F. Hoffstadt[a], Jason Nak[b], Jill de Ron[b], Koen Derks[l], Karoline Huth[a,m], Sjoerd Terpstra[n], Thomas Bastelica[o,p], Magda Matetovici[b,q], Vincent L. Ott[b], Andreea S. Zetea[b], Katharina Karnbach[b], Michelle C. Donzallaz[a], Arne John[a], Roy M. Moore[b], Franziska Assion[r], Riet van Bork[a], Theresa E. Leidinger[b], Xiaochang Zhao[b], Adrian Karami Motaghi[b], Ting Pan[b,i], Hannah Armstrong[b], Tianqi Peng[b,†], Mara Bialas[b,s], Joyce Y.-C. Pang[b], Bohan Fu[b], Shujun Yang[b], Xiaoyi Lin[b], Dana Sleiffer[b], Miklos Bognar[t], Balazs Aczel[u], and Eric-Jan Wagenmakers[a]

[a]Department of Psychological Methods, University of Amsterdam, Amsterdam, Netherlands; [b]Department of Psychology, University of Amsterdam, Amsterdam, Netherlands; [c]Institute of Economic Law, University of Kassel, Kassel, Germany; [d]Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium; [e]Chairs of Statistics and Econometrics, Georg August University of Göttingen, Göttingen, Germany; [f]Centre for Environmental Sciences, Hasselt University, Hasselt, Belgium; [g]Gutenberg School Wiesbaden, Wiesbaden, Germany; [h]Department of Psychology, Justus Liebig University, Giessen, Germany; [i]Department of Adult Psychiatry, Amsterdam University Medical Centre, Amsterdam, Netherlands; [j]Centre for Urban Mental Health, University of Amsterdam, Amsterdam, Netherlands; [k]Cognitive and Behavioral Decision Research, University of Zurich, Zurich, Switzerland; [l]Center of Accounting, Auditing & Control, Nyenrode Business University, Breukelen, Netherlands; [m]Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands; [n]Institute for Marine and Atmospheric Research Utrecht, Utrecht University, Utrecht, Netherlands; [o]Centre Hospitalier Le Vinatier, Bron, France; [p]INSERM, U1028; CNRS, UMR5292, Lyon Neuroscience Research Center, Psychiatric Disorders: from Resistance to Response Team, Lyon, France; [q]Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, Netherlands; [r]Faculty of Behavioral and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands; [s]Behavioural Science Institute, Radboud University, Nijmegen, Netherlands; [t]Doctoral School of Psychology, ELTE Eotvos Lorand University, Budapest, Hungary; [u]Institute of Psychology, ELTE Eotvos Lorand University, Budapest, Hungary

ABSTRACT

Many people have flipped coins but few have stopped to ponder the statistical and physical intricacies of the process. We collected 350,757 coin flips to test the counterintuitive prediction from a physics model of human coin tossing developed by Diaconis, Holmes, and Montgomery (DHM; 2007). The model asserts that when people flip an ordinary coin, it tends to land on the same side it started. Our data support this prediction: the coins landed on the same side more often than not, Pr(same side) $=$ 0.508, 95% credible interval (CI) [0.506, 0.509], $\text{BF}_{\text{same-side bias}} = 2359$. Furthermore, the data revealed considerable between-people variation in the degree of this same-side bias. Our data also confirmed the generic prediction that when people flip an ordinary coin—with the initial side-up randomly determined—it is equally likely to land heads or tails:Pr(heads) $=$ 0.500, 95% CI [0.498, 0.502], $\text{BF}_{\text{heads-tails bias}} = 0.182$. Additional analyses revealed that the within-people same-side bias decreased as more coins were flipped, an effect that is consistent with the possibility that practice makes people flip coins in a less wobbly fashion. Our data therefore provide strong evidence that when some (but not all) people flip a fair coin, it tends to land on the same side it started. Supplementary materials for this article are available online, including a standardized description of the materials available for reproducing the work.

ARTICLE HISTORY
Received September 2023
Accepted June 2025

KEYWORDS
Bayesian model-averaging; Chance; Informed hypothesis; Physics; Probability; Randomness

1. Introduction

A coin flip—the act of spinning a coin into the air with your thumb and then catching it in your hand—is often considered the epitome of a chance event. It features as a ubiquitous example in textbooks on probability theory and statistics (Kerrich 1946; Feller 1957; Gelman and Nolan 2002; Jaynes 2003; Küchenhoff 2008) and constituted a game of chance ("capita aut navia"–"heads or ships") already in Roman times (Macrobius, ∼ 431 AD, 1.7:22).

The simplicity and perceived fairness of a coin flip, coupled with the widespread availability of coins, may explain why it is

CONTACT František Bartoš f.bartos96@gmail.com Department of Psychological Methods, University of Amsterdam, Nieuwe Achtergracht 129B, Letter: PO Box 15906, 1001 NK Amsterdam, Parcel: Valckenierstraat 59, 1018 XE Amsterdam, Netherlands.
*Additional affiliation: Institute on Interdisciplinary Science, Berlin, Germany.
**Additional affiliation: Faculty of Psychology and Educational Sciences, Ghent University, Ghent, Belgium.
†Additional affiliation: Department of Organization and Personnel Management, Rotterdam School of Management, Erasmus University Rotterdam, Rotterdam, Netherlands.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

often used to make even high-stakes decisions. For example, a coin flip was used to determine which of the Wright brothers would attempt the first flight in 1903; who would get the last plane seat for the tour of rock star Buddy Holly (which crashed and left no survivors) in 1959; who would be the winner of the 1968 European Championship semifinal soccer match between Italy and the Soviet Union (an event which Italy went on to win); which of two companies would be awarded a public project in Toronto in 2003; to break the tie in local political elections in the Philippines in both 2004 and 2013; finally, this year a coin flip was used to determine whether a $5,000,000 game show prize would be doubled or whether the contestant would be eliminated.

Despite the widespread popularity of coin flipping, few people pause to reflect on the notion that the outcome of a coin flip is anything but random: a coin flip obeys the laws of Newtonian physics in a relatively transparent manner (Jaynes 2003). According to the standard model of coin flipping (Vulović and Prange 1986; Keller 1986; Engel 1992; Strzałko et al. 2008, 2010), the flip is a deterministic process and the perceived randomness originates from small fluctuations in the initial conditions (regarding starting position, configuration, upward force, and angular momentum) combined with narrow boundaries on the outcome space. Therefore, the standard model predicts that when people flip a fair coin, the probability of it landing heads is 50% (i.e., there is no "heads-tails bias"; conversely, if a coin would land on one side more often than the other, we would say there is a "heads-tails bias").[1]

The standard model of coin flipping was extended by Diaconis, Holmes, and Montgomery (DHM; 2007) who proposed that when people flip a ordinary coin, they introduce a small degree of "precession" or wobble—a change in the direction of the axis of rotation throughout the coin's trajectory. According to the DHM model, precession causes the coin to spend more time in the air with the initial side facing up. Consequently, the coin has a higher chance of landing on the same side as it started (i.e., "same-side bias"). Under the DHM model, this same-side bias is absent only when there is no wobble whatsoever, as any nonzero angle of rotation results in a same-side bias (with a higher degree of wobble resulting in a more pronounced bias). Based on a modest number of empirical observations (featuring coins with ribbons attached and high-frame-rate video recordings) Diaconis, Holmes, and Montgomery (2007) measured the off-axis rotations in typical human flips. Based on these observations, the DHM model predicted that a coin flip should land on the same side as it started with a probability of approximately 51%, just a fraction higher than chance.

Throughout history, several researchers have collected thousands of coin flips. In the 18th century, the famed naturalist Count de Buffon (1777) collected 2048 uninterrupted sequences of "heads" in what is possibly the first statistical experiment ever conducted. In the 19th century, the statistician Karl Pearson (1897) flipped a coin 24,000 times to obtain 12,012 tails. And in the 20th century, the mathematician John Kerrich (1946) flipped a coin 10,000 times for a total of 5067 heads. These experiments do not allow a test of the DHM model, however, mostly because it was not recorded whether the coin landed on the same side that it started. A notable exception is a sequence of 40,000 coin flips collected by Janet Larwood and Priscilla Ku (Berkeley 2009): Larwood always started the flips heads-up, and Ku always tails-up. Whereas Larwood's 10,231 out of 20,000 heads-to-heads flips are indicative of a same-side bias, Ku's 10,014 out of 20,000 tails-to-tails flips are not. Taken together, the 40,000 coin tosses were interpreted to "yield ambiguous evidence for dynamical bias" (Berkeley 2009).

In order to carry out a diagnostic empirical test of the same-side bias hypothesized by DHM, we collected a total of 350,757 coin flips, a number that—to the best of our knowledge—dwarfs all previous efforts. To anticipate our main results, the data reveal overwhelming statistical evidence for the presence of same-side bias. However, this effect needs to be qualified in the sense that it varies across individuals. Moreover, the bias appears to wane with practice. The data also yield moderate evidence for the complete absence of a heads-tails bias. The appendices demonstrate that the same conclusions obtain under a wide range of alternative analysis strategies.

2. Methods

We collected data in three different settings using the same standardized protocol. First, a group of five bachelor students collected at least 15,000 coin flips each as a part of their bachelor thesis project, contributing 75,036 coin flips in total. Second, we organized a series of on-site "coin flipping marathons" where 35 people spent up to 12 hr coin-flipping (see e.g., *https://www.youtube.com/watch?v=3xNg51mv-fk* for a video recording of one of the events), contributing a total of 203,440 coin flips.[2] Third, we issued a call for collaboration via Twitter, which resulted in an additional seven people contributing a total of 72,281 coin flips.

We encouraged people to flip coins of various currencies and denominations to ascertain the generalizability of the effect. Furthermore, we encouraged coin tossers to exchange coins, as this potentially allows people-specific effects to be disentangled from coin-specific effects. Overall, a group of 48 people[3] (i.e., all but three of the co-authors) tossed coins of 44 different currencies $\times$ denominations and obtained a total number of 350,757 coin flips.

The protocol required that each person collects sequences of 100 consecutive coin flips.[4] In each sequence, people randomly (or according to an algorithm) selected a starting position (heads-up or tails-up) of the first coin flip, flipped the coin, caught it in their hand, recorded the landing position of the coin (heads-up or tails-up), and proceeded with flipping the coin starting from the same side it landed in the previous trial

[1] Some even assert that a biased coin is a statistical unicorn—everyone talks about it but no one has actually encountered one (Gelman and Nolan 2002). Physics models support this assertion as long as the coin is not bent (Woo and Oh 2010) or allowed to spin on the ground (Jaynes 2003; Küchenhoff 2008).

[2] Including 2700 coin flips collected by the first two authors on a separate occasion.

[3] One of the bachelor students collected data with a family member who is counted as a "coin-tosser" but who did not qualify for co-authorship.

[4] Some sequences slightly varied in length due to issues with keeping track of the number of flips.

(we decided for this "autocorrelated" procedure as it simplified recording of the outcomes). In case the coin was not caught in hand, the flip was designated as a failure, and repeated from the same starting position. To simplify the recording and minimize coding errors, participants usually marked sides of the coins with permanent marker. To safeguard the integrity of the data collection effort, all participants videotaped and uploaded recordings of their coin flipping sequences.[5] See *https://osf.io/g8fzn/* for the data and video recordings.

3. Analysis

3.1. Same-Side Bias

An initial analysis confirms the prediction from the DHM model: the coins landed how they started more often than 50%. Specifically, the data feature 178,079 same-side landings out of 350,757 tosses, Pr(same side) $= 0.5077$, 95% central credible interval (CI) [0.5060, 0.5094] (under a binomial model with a uniform prior distribution), which is remarkably close to DHM's prediction of (approximately) 51%.

Evidence in favor of the DHM model's same-side bias prediction against the absence of the same-side bias can be evaluated using an informed Bayesian binomial test with k same-side outcomes out of N trials, $k \sim \text{Binomial}(\beta, N)$, assuming that the coin flips are independently and identically distributed across people and coins. We specified two competing hypotheses via the binomial success parameter β, where success denotes the coin landing on the same side it started from:

$$\begin{aligned} \text{No same-side bias, } \mathcal{H}_0 &: \beta = 0.5 \\ \text{DHM same-side bias, } \mathcal{H}_1 &: \beta \sim \text{Beta}(5100, 4900)_{[0.5,1]}. \end{aligned} \quad (1)$$

The highly informed $\text{Beta}(5100, 4900)_{[0.5,1]}$ prior distribution is meant to adequately represent the DHM hypothesis of the same-side bias (Diaconis, Holmes, and Montgomery 2007; see Appendix A for more details about the prior distribution settings and prior sensitivity analysis). The evidence is quantified by the Bayes factor (Jeffreys 1935, 1939; Kass and Raftery 1995; Etz and Wagenmakers 2017):

$$\text{BF}_{10} = \frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)},$$

which contrasts the competing hypotheses in terms of their predictive performance for the observed data. The Bayes factor hypothesis test indicates extreme evidence in favor of the same-side bias predicted by the DHM model, $\text{BF}_{\text{same-side bias}} = 1.76 \times 10^{17}$.

3.2. Heads-Tails Bias

Before proceeding to an analysis of heterogeneity in the same-side bias, we perform a similar analysis for the presence versus absence of the heads-tails bias, that is, $k \sim \text{Binomial}(\alpha, N)$ where k corresponds to the number of heads out N tosses with probability of landing on heads α. Specifically, we obtained 175,421 heads out of 350,757 tosses, Pr(heads) $= 0.5001$, 95% CI [0.4985, 0.5018]. Replacing the DHM same-side bias prior distribution with a highly informed Beta(5000, 5000) prior distribution (representing the background knowledge that in case a heads or tails bias exists, the effect is likely to be very small) yields moderate evidence against the presence of a heads-tails bias, $\text{BF}_{\text{heads-tails bias}} = 0.168$ (see Table 2 for by-coin summary).

3.3. Heterogeneity: Between-People and Between-Coins

A closer examination of the individual results, however, suggests that this overall result needs to be qualified in the sense that the size of the same-side bias varies across individuals (see Table 1). We extend the binomial model into a hierarchical model with by-person $k = 1, \ldots, K$ and by-coin $j = 1, \ldots, J$ specific deviations γ_{β_k} and γ_{α_j} from the overall same-side and the heads-tails bias (i.e., $\text{logit}(\beta_\mu)$ and $\text{logit}(\alpha_\mu)$, respectively[6]):

$$\begin{aligned} \gamma_{\alpha_j} &\sim \text{Normal}(0, \sigma_\alpha^2) \\ \gamma_{\beta_k} &\sim \text{Normal}(0, \sigma_\beta^2) \\ \text{logit}(\alpha_j) &= \underbrace{\text{logit}(\alpha_\mu)}_{\text{Overall same-side bias}} + \underbrace{\gamma_{\alpha_j}}_{\substack{\text{Person-specific deviation} \\ \text{from same-side bias}}} \\ \text{logit}(\beta_k) &= \underbrace{\text{logit}(\beta_\mu)}_{\text{Overall heads-tails bias}} + \underbrace{\gamma_{\beta_k}}_{\substack{\text{Coin-specific deviation} \\ \text{from heads-tails bias}}} \\ \mu_{ijk} &= \begin{cases} \text{logit}(\alpha_j) + \text{logit}(\beta_k) & y_{s=0,ijk} = 1 \text{ (Starting heads)} \\ \underbrace{\text{logit}(\alpha_j)}_{\text{Heads bias}} - \underbrace{\text{logit}(\beta_k)}_{\substack{\text{Same-side} \\ \text{bias}}} & y_{s=0,ijk} = 0 \text{ (Starting tails)} \end{cases} \\ y_{s=1,ijk} &\sim \text{Bernoulli}(\text{logit}^{-1}(\mu_{ijk})), \end{aligned} \quad (2)$$

where $y_{s=1,ijk} = 1$ corresponds to the ith flip of the kth person with the jth coin landing heads.[7] We first use the model with slightly informative prior distributions tailored for parameter estimation, and then proceed with more informed prior distributions tailored for hypothesis testing and Bayesian model-averaging to test simultaneously for the presence versus absence of the same-side bias, heads-tails bias, and between-people and between-coin heterogeneities in the respective biases (see, Jeffreys 1939, 1961). See Appendix A for more details.

3.3.1. Parameter Estimation

The observed proportion (and 95% CI) of same-side outcomes for each individual participant is shown as the blue dots (and bars) in Figure 1. It is clear that participants notably differ in the degree of the same-side bias. The black dots correspond to the estimated probability of the same-side outcome from

[5]There are occasional missing recordings due to failures of recording apparatus/lost files.

[6]While β_μ and α_μ correspond to the overall probability of the same-side and heads respectively, the logit() function transforms the overall probability into the bias (e.g., equal probability of heads and tails, $\alpha_\mu = 0.5$, corresponds to no heads-tails bias, $\text{logit}(\alpha_\mu) = 0$).

[7]The model could be further extended by including a coin-specific same-side bias and an associated person-coin same-side bias interaction. For the sake of simplicity, we do not pursue this model here but refer to Appendix C for results that suggest the absence of between-coin heterogeneity in same-side bias. The model also does not feature person-specific random-effects in heads-tails bias as we would not expect that some people are more likely to flip heads than tails or vice versa.

Table 1. By-person summary of the probability of a same side landing.

Person	Same side	Flips	Coins	Proportion [95% CI]	Joined
XiaoyiL	780	1600	2	0.487 [0.463, 0.512]	Marathon-MSc
JoyceP	1126	2300	3	0.490 [0.469, 0.510]	Marathon-MSc
AndreeaZ	2204	4477	4	0.492 [0.478, 0.507]	Marathon-MSc
KaleemU	7056	14324	8	0.493 [0.484, 0.501]	Bc Thesis
FelipeFV	4957	10015	3	0.495 [0.485, 0.505]	Internet
ArneJ	1937	3900	4	0.497 [0.481, 0.512]	Marathon-MSc
AmirS	7458	15012	6	0.497 [0.489, 0.505]	Bc Thesis
ChrisGl	4971	10005	5	0.497 [0.487, 0.507]	Marathon-Manheim
FrederikA	5219	10500	5	0.497 [0.487, 0.507]	Internet
FranziskaN	5368	10757	3	0.499 [0.490, 0.508]	Internet
JasonN	3352	6700	7	0.500 [0.488, 0.512]	Marathon-PhD
RietvanB	1801	3600	4	0.500 [0.484, 0.517]	Marathon-PhD
PierreG	7506	15000	9	0.500 [0.492, 0.508]	Bc Thesis
KarolineH	2761	5500	5	0.502 [0.489, 0.515]	Marathon-PhD
SjoerdT	2510	5000	5	0.502 [0.488, 0.516]	Marathon-MSc
SaraS	5022	10000	3	0.502 [0.492, 0.512]	Marathon-Manheim
HenrikG	8649	17182	8	0.503 [0.496, 0.511]	Marathon
IrmaT	353	701	1	0.504 [0.467, 0.540]	Bc Thesis
KatharinaK	2220	4400	5	0.504 [0.490, 0.519]	Marathon-PhD
JillR	3261	6463	2	0.505 [0.492, 0.517]	Marathon
FrantisekB	10,148	20,100	11	0.505 [0.498, 0.512]	Marathon
IngeborgR	4340	8596	1	0.505 [0.494, 0.515]	Marathon
VincentO	2475	4900	5	0.505 [0.491, 0.519]	Marathon-MSc
EricJW	2071	4100	5	0.505 [0.490, 0.520]	Marathon-MSc
MalteZ	5559	11,000	7	0.505 [0.496, 0.515]	Marathon-Manheim
TheresaL	1769	3500	4	0.505 [0.489, 0.522]	Marathon-MSc
DavidV	7586	14,999	5	0.506 [0.498, 0.514]	Bc Thesis
AntonZ	5069	10,004	2	0.507 [0.497, 0.516]	Marathon-Manheim
MagdaM	2510	4944	6	0.508 [0.494, 0.522]	Marathon-MSc
ThomasB	2540	5000	5	0.508 [0.494, 0.522]	Marathon-PhD
JonasP	5080	9996	5	0.508 [0.498, 0.518]	Marathon
BohanF	1118	2200	3	0.508 [0.487, 0.529]	Marathon-MSc
HannahA	1525	3000	4	0.508 [0.490, 0.526]	Marathon-MSc
AdrianK	1749	3400	3	0.514 [0.498, 0.531]	Marathon-MSc
AaronL	3815	7400	5	0.515 [0.504, 0.527]	Marathon-MSc
KoenD	3309	6400	7	0.517 [0.505, 0.529]	Marathon-PhD
MichelleD	2224	4300	5	0.517 [0.502, 0.532]	Marathon-PhD
RoyMM	2020	3900	4	0.518 [0.502, 0.534]	Marathon-MSc
TingP	1658	3200	4	0.518 [0.501, 0.535]	Marathon-MSc
MaraB	1426	2750	3	0.518 [0.500, 0.537]	Marathon-MSc
AdamF	4335	8328	2	0.520 [0.510, 0.531]	Marathon
AlexandraS	9080	17,434	8	0.521 [0.513, 0.528]	Marathon
MadlenH	3705	7098	1	0.522 [0.510, 0.534]	Marathon
DavidKL	7895	15,000	1	0.526 [0.518, 0.534]	Bc Thesis
XiaochangZ	1869	3481	4	0.537 [0.520, 0.553]	Marathon-MSc
FranziskaA	2055	3800	4	0.541 [0.525, 0.557]	Marathon-MSc
JanY	956	1691	2	0.565 [0.542, 0.589]	Marathon-MSc
TianqiP	1682	2800	3	0.601 [0.582, 0.619]	Marathon-MSc
Combined	178,079	350,757	44	0.508 [0.506, 0.509]	

NOTE: "Proportion" refers to the observed proportion of coin flips that landed on the same side with a 95% central credible interval under uniform prior distributions (virtually identical to a frequentist confidence interval).

the hierarchical model with slightly informative prior distributions that incorporates both between-people heterogeneity in same-side bias and between-coin heterogeneity in heads-tails bias. Only a small proportion of participants' point estimates fall below the chance line (i.e., no bias), with the majority of point estimates somewhere in between the chance and the DHM model prediction, and some exceeding the DHM prediction. The left bottom panel of Figure 1 displays the posterior distribution of the probability of a same side outcome, Pr(same side) = 0.5098, 95% CI [0.5050, 0.5147]. This credible interval is slightly wider than the one from the simple binomial analysis; this is due to the substantial between-people heterogeneity in the probability of the coin landing on the same side, $\text{sd}_{\text{people}}(\text{Pr(same side)}) = 0.0156$, 95% CI [0.0119, 0.0200] (right bottom panels of Figure 1).

Inspection of the coin-specific parameters (Figure 2) suggests the lack of a heads-tails bias, Pr(heads) = 0.5005, 95% CI [0.4986, 0.5026], with virtually no between-coin heterogeneity, $\text{sd}_{\text{coins}}(\text{Pr(heads)}) = 0.0018$, 95% CI [0.0001, 0.0047], as all observed proportions and estimated probabilities of landing on heads cluster around the chance line.

3.3.2. Hypothesis Testing

Evidence in favor of the presence versus absence of the same-side bias, heads-tails bias, and between-people and between-coin heterogeneities can be evaluated simultaneously using Bayesian model-averaging (Raftery, Madigan, and Volinsky 1995; Hoeting et al. 1999; Hinne et al. 2020) and *inclusion Bayes factors*, a generalization of Bayes factors based on the change from prior to posterior odds (Hinne et al. 2020):

Table 2. By-coin summary of the probability of heads.

Coin	Heads	Flips	People	Proportion [95% CI]
0.25CAD	48	100	1	0.480 [0.379, 0.582]
20DEM-silver	484	1000	1	0.484 [0.453, 0.515]
5CZK	1222	2500	2	0.489 [0.469, 0.509]
0.05NZD	984	2011	1	0.489 [0.467, 0.511]
0.10EUR	4515	9165	6	0.493 [0.482, 0.503]
1DEM	2464	5000	5	0.493 [0.479, 0.507]
50CZK	3207	6500	7	0.493 [0.481, 0.506]
2HRK	4258	8596	1	0.495 [0.485, 0.506]
1MXN	4180	8434	1	0.496 [0.485, 0.506]
1SGD	7655	15,400	2	0.497 [0.489, 0.505]
5ZAR	3645	7325	1	0.498 [0.486, 0.509]
2EUR	24,276	48,772	28	0.498 [0.493, 0.502]
0.01GBP	498	1000	1	0.498 [0.467, 0.529]
0.50EUR	28,617	57,445	32	0.498 [0.494, 0.502]
0.20EUR	15,665	31,373	20	0.499 [0.494, 0.505]
0.25BRL	1998	4000	2	0.499 [0.484, 0.515]
0.10RON	1000	2001	1	0.500 [0.478, 0.522]
1CHF	2249	4500	4	0.500 [0.485, 0.514]
1EUR	18,920	37,829	25	0.500 [0.495, 0.505]
0.20GEL	4501	8998	5	0.500 [0.490, 0.511]
1CAD	5604	11,200	11	0.500 [0.491, 0.510]
2CAD	1502	3000	3	0.501 [0.483, 0.519]
2MAD	1503	3000	1	0.501 [0.483, 0.519]
100JPY	752	1500	1	0.501 [0.476, 0.527]
2CHF	2259	4503	2	0.502 [0.487, 0.516]
5MAD	1007	2001	1	0.503 [0.481, 0.525]
0.20GBP	1516	3005	2	0.504 [0.486, 0.523]
1CNY	757	1500	1	0.505 [0.479, 0.530]
1CZK	505	1000	1	0.505 [0.474, 0.536]
2ILS	506	1000	1	0.506 [0.475, 0.537]
5JPY	1772	3500	2	0.506 [0.490, 0.523]
5SEK	8052	15,902	7	0.506 [0.499, 0.514]
0.25USD	2180	4300	4	0.507 [0.492, 0.522]
1MAD	1014	2000	1	0.507 [0.485, 0.529]
0.50RON	1442	2844	3	0.507 [0.488, 0.526]
0.05EUR	3821	7514	6	0.509 [0.497, 0.520]
0.50GBP	765	1504	1	0.509 [0.483, 0.534]
2BDT	2038	4003	2	0.509 [0.494, 0.525]
10CZK	4572	8905	7	0.513 [0.503, 0.524]
0.20CHF	518	1000	1	0.518 [0.487, 0.549]
0.50SGD	1449	2781	3	0.521 [0.502, 0.540]
0.02EUR	158	300	1	0.527 [0.468, 0.584]
1GBP	791	1500	2	0.527 [0.502, 0.553]
2INR	552	1046	1	0.528 [0.497, 0.558]
Combined	175,421	350,757	48	0.500 [0.498, 0.502]

NOTE: "Proportion" refers to the observed proportion of coin flips that landed heads with a 95% central credible interval under uniform prior distributions (virtually identical to a frequentist confidence interval).

$$\underbrace{\text{BF}_{AB}}_{\substack{\text{Inclusion Bayes factor}\\ \text{for A vs. B}}} = \underbrace{\frac{\sum_{a\in A} p(\mathcal{M}_a \mid \text{data})}{\sum_{b\in B} p(\mathcal{M}_b \mid \text{data})}}_{\substack{\text{Posterior inclusion odds}\\ \text{for A vs. B}}} \Big/ \underbrace{\frac{\sum_{a\in A} p(\mathcal{M}_a)}{\sum_{b\in B} p(\mathcal{M}_b)}}_{\substack{\text{Prior inclusion odds}\\ \text{for A vs. B}}}, \tag{3}$$

where A contains a set of models where a given hypothesis holds and B contains the complement. The overarching set of all models (A and B) can be specified as an orthogonal combination of the different possible hypotheses, that is, $2^4 = 16$ models, where each is assigned the usual equal prior model probability of $1/16$ (Appendix A lists all specified models in detail).

The inclusion Bayes factor indicated compelling evidence for the presence of an overall same-side bias, $\text{BF}_{\text{same-side bias}} = 2359$. When the hypothesis of a same-side bias has a prior probability of 0.50, a Bayes factor of about 2359 results in a posterior probability of 0.9996. In addition, consistent with the visual impression from Figure 1, the inclusion Bayes factor reveals overwhelming evidence for the presence of between-people heterogeneity in same-side bias, $\text{BF}_{\text{people heterogeneity}} = 3.10 \times 10^{24}$. The evidence against the presence of heads-tails bias remains practically unchanged, $\text{BF}_{\text{heads-tails bias}} = 0.182$. The model indicates moderate evidence against the presence of between-coin heterogeneity in heads-tails bias, $\text{BF}_{\text{coin heterogeneity}} = 0.178$ (see Appendix C for concordant frequentist result and Table 1 in the Appendix H for an overview of the individual models).

3.4. Practice Effects

Following up on a suggestion by a reviewer, a more extensive inspection of the data suggested that the degree of same-side bias changes with the number of coin flips performed (i.e., practice) by each person. Moreover, the pattern and degree of change varied across participants. The top panel of Figure 3 shows the proportion of same-side outcomes aggregated by 1000 coin flips as black dots (and bars) for three participants with the largest number of coin flips. Each panel displays a qualitatively different pattern of change; the left panel shows a wavy pattern of same-side bias with practice, the middle panel shows a pattern in which the same-side bias first decreases and then remains stationary, and the right panel shows a pattern in which the same-side bias is first stationary and then decreases. While not all individual patterns fit into one of these three "pattern types", most participants showed an initial decrease in same-side bias with practice (see Appendix F for visualization of all participant trajectories). The bottom panel of Figure 3 shows the proportion of same-side outcomes aggregated by 1000 coin flips as black dots (and bars) combined across all participants. The overall pattern of same-side bias seems to be monotonically decreasing with practice. Note that the uncertainty increases at the higher numbers of coin flips, because only a few participants flipped coins more than 15,000 times.

The black trend lines and credible bands in Figure 3 correspond to the estimated probability of a same-side outcome from a hierarchical state-space model that monitors the change in the same-side bias with practice. State-space models are flexible nonparametric models often used for monitoring non-monotonic and nonlinear changes over time (Frühwirth-Schnatter 2006). Here, the hierarchical state-space model extends the hierarchical binomial model (2) by modeling the current state of the same-side bias of the ith flip of the kth person as a person-and-flip-specific change δ_{ki} from their previous state of the same-side bias, $\beta_{k(i-1)}$, and a constant person-specific drift parameter θ_k,

$$\begin{aligned}
\gamma_{\beta_k} &\sim \text{Normal}(0, \sigma^2_\beta) \\
\gamma_{\delta_{ki}} &\sim \text{Normal}(0, \sigma^2_\delta) \\
\gamma_{\theta_k} &\sim \text{Normal}(0, \sigma^2_\theta) \\
\delta_{ki} &= \underbrace{\delta_{\mu i}}_{\substack{\text{Overall state change}\\ \text{at flip } i}} + \underbrace{\gamma_{\delta_{ki}}}_{\substack{\text{Person-specific deviation}\\ \text{from state change at flip } i}} \\
\theta_k &= \underbrace{\theta_\mu}_{\text{Overall drift}} + \underbrace{\gamma_{\theta_k}}_{\substack{\text{Person-specific deviation}\\ \text{from drift}}}
\end{aligned} \tag{4}$$

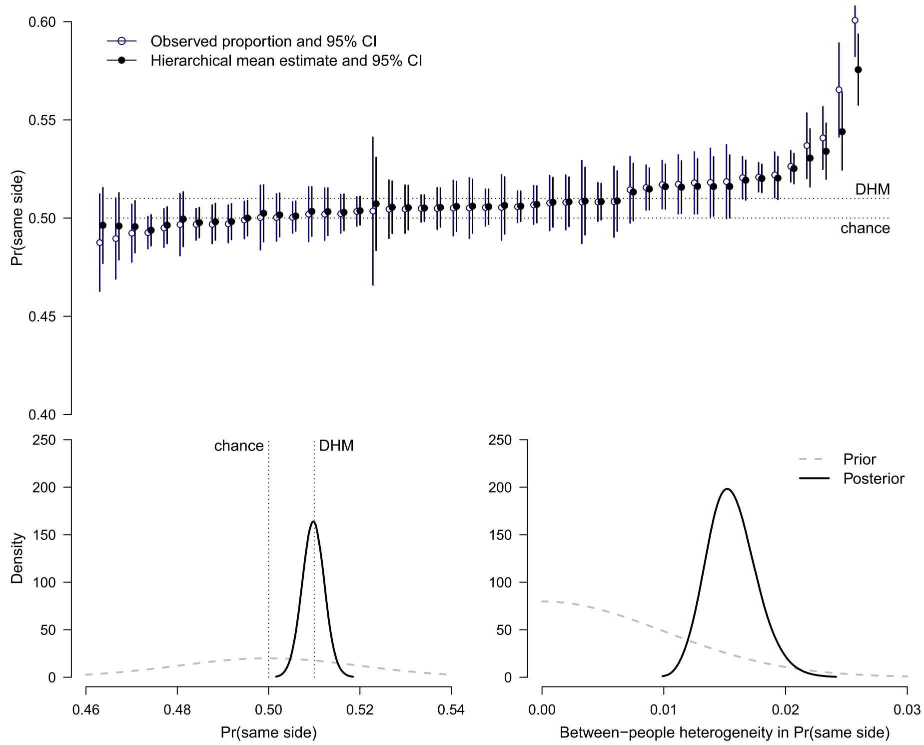

Figure 1. Coins have a tendency to land on the same side they started, confirming the predictions from the Diaconis, Holmes, and Montgomery (DHM) model of coin flipping. Top panel: posterior estimates of the probability of same side separately for each person, as obtained from the hierarchical Bayesian model with weakly informative, estimation-tailored prior distributions described in the methods section; Bottom-left panel: prior and posterior distributions for the overall probability of same side; Bottom-right panel: prior and posterior distributions for the between-people heterogeneity in the probability of the same side.

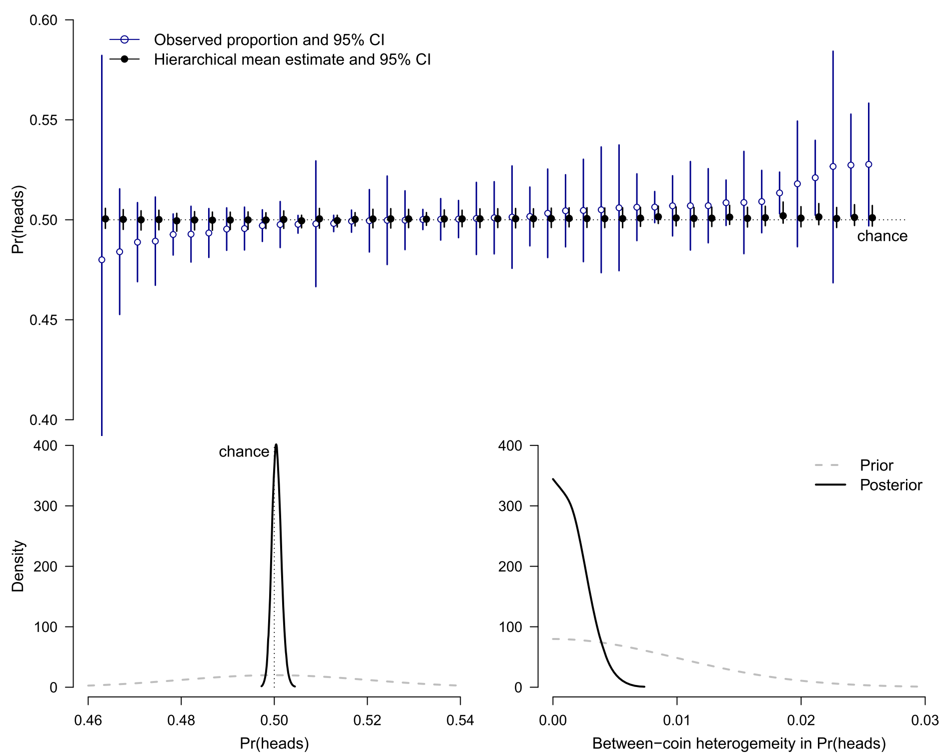

Figure 2. Coins have a tendency to land on heads and tails with equal probability, supporting the predictions from the Standard model of coin flipping. Top panel: posterior estimates of the probability of heads separately for each coin, as obtained from the hierarchical Bayesian model with weakly informative, estimation-tailored prior distributions described in the methods section; Bottom-left panel: prior and posterior distributions for the overall probability of heads; Bottom-right panel: prior and posterior distributions for the between-coin heterogeneity in the probability of heads.

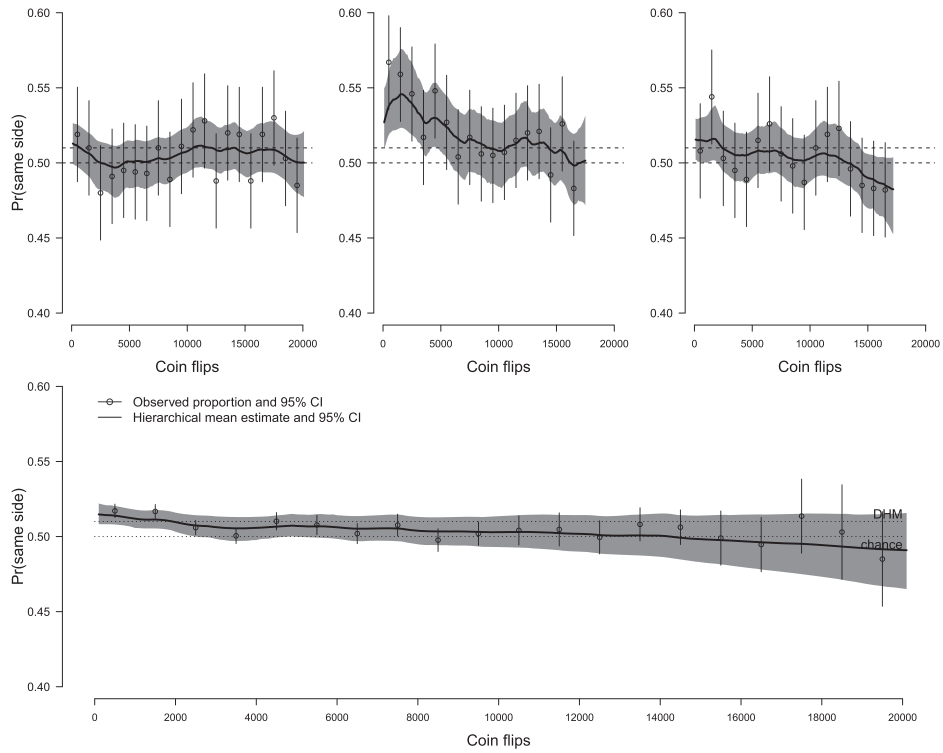

Figure 3. The probability of coins landing on the same side changes with practice. Top panel: posterior estimates of the probability of the same side separately for three participants with the largest number of coin flips across the number of conducted coin flips (see Appendix F for all trajectories); Bottom panel: posterior estimates of the probability of the same side across all participants showing on average decreasing pattern of the probability of the same side across the number of conducted coin flips. The posterior estimates are obtained from the hierarchical Bayesian state-space model with weakly informative, estimation-tailored prior distributions described in the methods section overlaying the observed proportions binned by 1000 coin flips.

$$\underbrace{\text{logit}(\beta_{ki})}_{\substack{\text{Current state}\\ \text{of same-side bias}}} = \begin{cases} \underbrace{\text{logit}(\beta_{\mu 1})}_{\substack{\text{Overall initial}\\ \text{same-side bias}}} + \underbrace{\gamma_{\beta_{k1}}}_{\substack{\text{Initial person-specific deviation}\\ \text{from same-side bias}}} \\ i = 1 \text{ (Initial state)} \\ \underbrace{\text{logit}(\beta_{k(i-1)})}_{\substack{\text{Previous state}\\ \text{of same-side bias}}} + \underbrace{\delta_{ki}}_{\substack{\text{State change}\\ \text{at flip } i}} + \underbrace{\theta_k}_{\text{Drift}} \\ i > 1 \text{ (Subsequent states)} \end{cases}$$

$$\mu_{ijk} = \begin{cases} \text{logit}(\beta_{ki}) & y_{s=0,ijk} = 1 \text{ (Starting heads)} \\ -\text{logit}(\beta_{ki}) & y_{s=0,ijk} = 0 \text{ (Starting tails)} \end{cases}$$

$$y_{s=1,ijk} \sim \text{Bernoulli}(\text{logit}^{-1}(\mu_{ijk})),$$

where $\delta_{\mu i}$ corresponds to the overall same-side bias at ith flip and θ_μ corresponds to the overall drift in the same-side bias. In contrast to the earlier model, (2), we omit the overall heads-tails bias and the between-coin heterogeneity in the heads-tails bias to simplify the model estimation process in light of the evidence against the presence of the heads-tails bias and its heterogeneity. We only use the model with slightly informative prior distributions tailored for parameter estimations, see Appendix A for more details.

The pooled estimate of the initial same-side bias based on the state-space model equals Pr(same side) $=$ 0.5148, 95% CI [0.5082, 0.5220] and is larger than the overall same-side bias estimate from the hierarchical model. However, the initial heterogeneity in the probability of the coin landing on the same side (based on the state-space model; $\text{sd}_{\text{people}}(\text{Pr(same side)}) =$ 0.0082, 95% CI [0.0008, 0.0162]) is about half the overall heterogeneity in the probability of coins landing on the same side (based on the hierarchical model; $\text{sd}_{\text{people}}(\text{Pr(same side)}) =$ 0.0156, 95% CI [0.0119, 0.0200]). This result suggests the large between-people heterogeneity observed in the hierarchical model was confounded to some degree by practice effects.

3.5. Outlier Exclusion

We repeated the statistical analyses after excluding four potential outliers with same-side sample proportions larger than 53% (i.e., the four largest and right-most estimates in the top panel of Figure 1). In general, the exclusion did not qualitatively affect our conclusions, although—as may be expected—the same-side bias decreased in size, and the between-people heterogeneity became less pronounced. Each of the four participants with the same-side sample proportion larger than 53% contributed fewer than 4000 coin flips, which is consistent with the possibility that the relatively high proportions may be due in part to the fact that the same-side bias is largest at the beginning (before the practice effects occur). Additional robustness checks reported in Tables 3–5 in the Appendix H demonstrate that the qualitative conclusions do not change when excluding 1 to 5 participants with the lowest and the largest proportion of same-side outcomes.

After excluding the four potential outliers, the data feature 171,517 same-side landings from 338,985 tosses, Pr(same side) $=$ 0.5060, 95% CI [0.5043, 0.5077]. The evidence in favor of the DHM hypothesis using the Bayesian informed binomial hypothesis test decreased notably but remains extreme, $BF_{\text{same-side bias}} = 1.28 \times 10^8$. The proportion of heads remained practically identical, 169,635 heads out of 338,985 tosses, Pr(heads) = 0.5004, 95% CI [0.4987, 0.5021], as did the moderate evidence against heads-tails bias $BF_{\text{heads-tails bias}}$ = 0.190.

The exclusion of potential outliers similarly affects the inference from the hierarchical model: although the same-side bias decreases and the associated heterogeneity is reduced (i.e., Pr(same side) = 0.5060, 95% CI [0.5031, 0.5089] and sd_{people}(Pr(same side)) = 0.0072, 95% CI [0.0050, 0.0099]), the evidence for the presence of the same-side bias and the associated between-people heterogeneity remains extreme (i.e., $BF_{\text{same-side bias}}$ = 787 and $BF_{\text{people heterogeneity}} = 2.87 \times 10^7$). For the probability of heads, the inference remains practically unchanged, both with respect to the size of the effect and associated heterogeneity (i.e., Pr(heads) = 0.5008, 95% CI [0.4988, 0.5030] and sd_{coins}(Pr(heads)) = 0.0020, 95% CI [0.0001, 0.0050]), and with respect to the evidence for the presence of the heads-tails bias and the associated between-coin heterogeneity (i.e., $BF_{\text{heads-tails bias}} = 0.213$ and $BF_{\text{coin heterogeneity}} = 0.221$).

The exclusion of potential outliers affects the practice effects model in a similar manner. The initial same-side bias slightly decreased, Pr(toss-order same side) = 0.5104, 95% CI [0.5049, 0.5167], and the associated between-people heterogeneity was reduced, sd_{people}(Pr(toss-order same side)) = 0.0036, 95% CI [0.0003, 0.0099].

3.6. *Including the Results from Larwood and Ku*

We repeated the statistical analyses after including the 40,000 coin flips performed by Larwood and Ku Berkeley (2009). In general, the inclusion did not qualitatively affect our conclusion, although the evidence for the same-side bias slightly increased.

After inclusion of the 40,000 coin flips, the data feature 198,324 same-side landings from 390,757 tosses, Pr(same side) = 0.5075, 95% CI [0.5060, 0.5091]. The evidence in favor of the DHM hypothesis using the Bayesian informed binomial hypothesis test slightly increases, $BF_{\text{same-side bias}} = 2.77 \times 10^{18}$. The proportion of heads remained practically identical, 195,638 heads out of 390,757 tosses, Pr(heads) = 0.5007, 95% CI [0.4991, 0.5022], as did the moderate evidence against heads-tails bias $BF_{\text{heads-tails bias}} = 0.221$.

The inclusion of the 40,000 coin flips affects the inference from the hierarchical model in a similar fashion: the degree of the same-side bias and the associated heterogeneity remains practically unchanged (i.e., Pr(same side) = 0.5096, 95% CI [0.5050, 0.5142] and sd_{people}(Pr(same side)) = 0.0152, 95% CI [0.0116, 0.0194]), the evidence for the presence of the same-side bias increases, and the evidence for the associated between-people heterogeneity slightly decreases (i.e., $BF_{\text{same-side bias}}$ = 3295 and $BF_{\text{people heterogeneity}} = 1.66 \times 10^{24}$). For the probability of heads, the inference also remains practically unchanged, both with respect to the size of the effect and associated heterogeneity (i.e., Pr(heads) = 0.5005, 95% CI [0.4986, 0.5026] and sd_{coins}(Pr(heads)) = 0.0017, 95% CI [0.0001, 0.0047]), and with respect to the evidence for the presence of the heads-tails bias and the associated between-coin heterogeneity (i.e., $BF_{\text{heads-tails bias}} = 0.183$ and $BF_{\text{coin heterogeneity}} = 0.175$).

The inclusion of the 40,000 coin flips similarly affects the practice effects model. The initial same-side bias slightly decreased, Pr(toss-order same side) = 0.5132, 95% CI [0.5074, 0.5195], and the associated between-people heterogeneity was reduced, sd_{people}(Pr(toss-order same side)) = 0.0071, 95% CI [0.0006, 0.0148].

4. Discussion

We collected 350,757 coin flips and found strong empirical evidence for the counterintuitive and precise prediction from DHM model of human coin tossing: when people flip a coin, it tends to land on the same side as it started. However, this conclusion needs to be qualified by two important factors. First, the data revealed a substantial degree of between-people variability in the same-side bias: as can be seen in Figure 1, some people appear to have little or no same-side bias, whereas others do display a same-side bias, albeit to a varying degree. This variability is arguably consistent with DHM model, in which the same-side bias originates from off-axis rotations (i.e., precession or wobbliness), which can reasonably be assumed to vary between people. Second, additional exploratory analyses suggest that the degree of the same-side bias decreases with the number of coin flips. A possible explanation of this decreasing bias is a coin-tossing practice effect—the more coins people flip, the closer they approach the "perfect" wobble-less flip. Our results suggest that around 10,000 coin flips ($\approx$ 10 hr of coin flipping) might be enough to virtually eliminate the same-side bias. Our results are robust to the type of coins used and to changes in analytic methodology—for instance, different prior distributions as outlined in Appendix B, different analytical approaches as outlined in Appendix D and discussed by Viechtbauer (2024) and Pawel (2024), and different criteria for outlier exclusion as outlined in the Appendix H. However, our experiment cannot rule out alternative explanations of the reduction in the same-side bias over time, such as exhaustion, waning attention, etc.

Our results are aligned with previous empirical data evaluating same-side bias in coin flipping; the Diaconis, Holmes, and Montgomery (2007) 27 high-speed camera flips that initially suggested the same-side bias of approximately 1%, and the subsequent 20,000 coin flips by Janet Larwood and Priscilla Ku that resulted in a same-side bias of 1.2% and 0.1%, respectively (results that incidentally also highlight the between-people heterogeneity in the same-side bias; Berkeley 2009). Furthermore, in a recent informal replication, McGaw (2024) collected 10 coin flips from 820 people each, resulting in 4171 same-side outcomes, for a same-side proportion of 0.5087.

Future work may attempt to verify whether "wobbly tossers" show a more pronounced same-side bias than "stable tossers" and examine the practice effects in more detail. Furthermore, the present practice effects suggest that the same-side bias is best studied through many people who each contribute only a few thousand coin flips rather than through few people who each contribute tens of thousands of coin flips; the same-side bias is much more pronounced at the beginning of the experiment, and hence there are diminishing returns when a single person contributes very many coin flips. However, the effort required to test the more detailed hypotheses appears to be excessive, as this would ideally involve detailed analyses of high-speed camera recordings for individual flips (see Diaconis, Holmes, and Montgomery 2007).

In order to ensure the quality of the data, we videotaped and audited the data collection procedure (see the Appendix E for details). The audit did not reveal anything suspicious (i.e., all participants performed the coin flips they reported). While the video recordings were of insufficient quality to provide additional insights into the variability of the same-side bias, McGaw (2024) noticed that the coin flipping technique of participants with the extreme same-side bias was visibly worse than that of the remaining participants. Exclusion of those participants, however, did not significantly affect the results. There also remains a legitimate concern: at the time when people were flipping the coins they were aware of the main hypothesis under test. Therefore, it cannot be excluded that some of the participants were able to manipulate the coin flip outcomes in order to produce the same-side bias. In light of the nature of the coin tossing process, the evidence from the video recordings, and the precise correspondence between the data and the predictions from DHM model, we deem this possibility unlikely, but future work is needed to disprove it conclusively (e.g., by concealing the aim of the study).

Could future coin tossers use the same-side bias to their advantage? The magnitude of the observed bias can be illustrated using a betting scenario. If you bet a dollar on the outcome of a coin toss (i.e., paying 1 dollar to enter, and winning either 0 or 2 dollars depending on the outcome) and repeat the bet 1000 times, knowing the starting position of the coin toss would earn you 16 dollars on average. This is more than the casino advantage for 6 deck blackjack against an optimal-strategy player, where the casino would make 5 dollars on a comparable bet, but less than the casino advantage for single-zero roulette, where the casino would make 27 dollars on average (Hannum 2003). Moreover, this advantage amounts to a non-negligible 95,000$ in expected value when considering the 5,000,000$ double-or-eliminated coin flip in the "Beast Games" Amazon Prime show (Conklin, Klitzner, and Beast 2025). These considerations lead us to suggest that when coin flips are used for high-stakes decision-making, the starting position of the coin is best concealed.

Supplementary Materials

The supplementary materials contain the following appendixes: Appendix A: Prior distributions. Appendix B: Prior sensitivity analysis. Appendix C: Same-side bias variability by recruitment site. Appendix D: Frequentist results Appendix E: Audit. Appendix F: Individual patterns of practice effects. Appendix G: Parametric practice effects. Appendix H: Additional tables and figures.

Author note

During a revision we corrected two minor errors in the dataset: (a) two coins were miscoded as different currency × denomination, and (b) one invalid coin flip with unknown starting position was not removed from the analysis as incorrectly treated as starting heads and landing tails. We fixed the coin coding issue, which reduced the number of unique coins from 46 to 44. We fixed the mistake in data processing and asked the participant to provide one additional valid coin flip to retain the same number of total coin flips. The additional flip started tails and landed tails, which increased the number of coins landing on the same side by one and decreased the number of coins landing on heads by one (we could not retrieve the original coin; consequently, the number of flips with 5 ZAR reduced by one and number of flips with 0.05 EUR increased by one). Our original findings were qualitatively unaffected by these errors.

Disclosure Statement

The authors have no competing interest to declare.

Funding

The authors have no funding to declare, and conducted this research in their spare time. For all but the last four positions, the authorship order aligns with the number of coin flips contributed.

Data Availability Statement

All data and materials are available at *https://osf.io/pxu6r/*.

Ethical Approval

The research project was approved by the Ethics Review Board of the Faculty of Social and Behavioral Sciences, University of Amsterdam, The Netherlands (2022-PML-15687).

ORCID

František Bartoš http://orcid.org/0000-0002-0018-5573
Alexandra Sarafoglou http://orcid.org/0000-0003-0031-685X
Henrik R. Godmann http://orcid.org/0009-0003-6410-4765
Pierre Y. Gui http://orcid.org/0009-0007-8065-597X
Malte Zoubek http://orcid.org/0000-0002-6944-5564
Frederik Aust http://orcid.org/0000-0003-4900-788X
Felipe Fontana Vieira http://orcid.org/0009-0006-0949-6569
Chris-Gabriel Islam http://orcid.org/0000-0003-4226-8884
Jonas Petter http://orcid.org/0009-0009-6609-5608
Ingeborg B. Roos http://orcid.org/0009-0001-9802-7644
Adam Finnemann http://orcid.org/0000-0003-4969-240X
Aaron B. Lob http://orcid.org/0000-0003-0459-7986
Madlen F. Hoffstadt http://orcid.org/0009-0001-3534-4335
Jason Nak http://orcid.org/0009-0009-5050-9917
Jill de Ron http://orcid.org/0000-0001-7226-3836
Koen Derks http://orcid.org/0000-0002-5533-9349
Karoline Huth http://orcid.org/0000-0002-0662-1591
Sjoerd Terpstra http://orcid.org/0000-0002-4580-0381
Thomas Bastelica http://orcid.org/0000-0003-1515-1488
Magda Matetovici http://orcid.org/0009-0002-3363-4037
Vincent L. Ott http://orcid.org/0000-0002-1325-8006
Katharina Karnbach http://orcid.org/0000-0002-6900-614X
Michelle C. Donzallaz http://orcid.org/0000-0001-6124-5528
Riet van Bork http://orcid.org/0000-0002-4772-8862
Adrian Karami Motaghi http://orcid.org/0009-0005-6599-0169
Hannah Armstrong http://orcid.org/0009-0009-9870-2978
Tianqi Peng http://orcid.org/0009-0002-4200-2867
Mara Bialas http://orcid.org/0000-0002-0533-1592
Joyce Y.-C. Pang http://orcid.org/0009-0007-4095-4676
Miklos Bognar http://orcid.org/0000-0003-3303-9852
Balazs Aczel http://orcid.org/0000-0001-9364-4988
Eric-Jan Wagenmakers http://orcid.org/0000-0003-1596-1034

References

Berkeley. (2009), "40,000 Coin Tosses Yield Ambiguous Evidence for Dynamical Bias," available at *https://www.stat.berkeley.edu/~aldous/Real-World/coin_tosses.html*. [2,8]

Buffon, G.-L. (1777), "Essai d'arithmétique morale," *Supplément à l'Histoire Naturelle*, 4, article XXIII. [2]

Conklin, T., Klitzner, S., and Beast, M. (2025), "Beast Games: Episode 10." [9]

Diaconis, P., Holmes, S., and Montgomery, R. (2007), "Dynamical Bias in the Coin Toss," *SIAM Review*, 49, 211–235. [2,3,8,9]

Engel, E. (1992), *A Road to Randomness in Physical Systems* (Vol. 71), New York: Springer-Verlag. [2]

Etz, A., and Wagenmakers, E.-J. (2017), "J. B. S. Haldane's Contribution to the Bayes Factor Hypothesis Test," *Statistical Science*, 32, 313–329. [3]

Feller, W. (1957), *An Introduction to Probability: Theory and Examples*, New York: Wiley. [1]

Frühwirth-Schnatter, S. (2006), *Finite Mixture and Markov Switching Models*, New York: Springer. [5]

Gelman, A., and Nolan, D. (2002), "You Can Load a Die, But You Can't Bias a Coin," *The American Statistician*, 56, 308–311. [1,2]

Hannum, R. C. (2003), "A Guide to Casino Mathematics," *UNLV Gaming Studies Research Center*. [9]

Hinne, M., Gronau, Q. F., van den Bergh, D., and Wagenmakers, E.-J. (2020), "A Conceptual Introduction to Bayesian Model Averaging," *Advances in Methods and Practices in Psychological Science*, 3, 200–215. [4]

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999), "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14, 382–401. [4]

Jaynes, E. T. (2003), *Probability Theory: The Logic of Science*, Cambridge: Cambridge University Press. [1,2]

Jeffreys, H. (1935), "Some Tests of Significance, Treated by the Theory of Probability," *Proceedings of the Cambridge Philosophy Society*, 31, 203–222. [3]

——— (1939), *Theory of Probability* (1st ed.), Oxford: Oxford University Press. [3]

——— (1961), *Theory of Probability* (3rd ed.), Oxford: Oxford University Press. [3]

Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795. [3]

Keller, J. B. (1986), "The Probability of Heads," *The American Mathematical Monthly*, 93, 191–197. [2]

Kerrich, J. E. (1946), *An Experimental Introduction to the Theory of Probability*, Copenhagen: E. Munksgaard. [1,2]

Küchenhoff, H. (2008), "Coin Tossing and Spinning–Useful Classroom Experiments for Teaching Statistics," in *Recent Advances in Linear Models and Related Areas*, pp. 417–426, Heidelberg: Physica-Verlag. [1,2]

Macrobius, A. A. T. ($\approx$ 431AD), Saturnalia. [1]

McGaw, A. (2024), "Personal Communication," Personal communication on 28 November 2024. Data from experiment accompanying available at *https://www.youtube.com/watch?v=-QjgvbvFoQA&ab_channel=AnotherRoof*. [8,9]

Pawel, S. (2024), "A Bayes Factor Framework for Unified Parameter Estimation and Hypothesis Testing," arXiv preprint. [8]

Pearson, K. (1897), *The Chances of Death, and Other Studies in Evolution*, London: Edward Arnold. [2]

Raftery, A. E., Madigan, D., and Volinsky, C. T. (1995), "Accounting for Model Uncertainty in Survival Analysis Improves Predictive Performance," in *Bayesian Statistics*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, pp. 323–349, Oxford: Oxford University Press. [4]

Strzałko, J., Grabski, J., Stefański, A., Perlikowski, P., and Kapitaniak, T. (2008), "Dynamics of Coin Tossing is Predictable," *Physics Reports*, 469, 59–92. [2]

——— (2010), "Understanding Coin-Tossing," *The Mathematical Intelligencer*, 32, 54–58. [2]

Viechtbauer, W. (2024), "Results of 350,757 Coin Flips to Examine Same-Side Bias." [8]

Vulović, V. Z., and Prange, R. E. (1986), "Randomness of a True Coin Toss," *Physical Review A*, 33, 576–582. [2]

Woo, D.-K., and Oh, C.-H. (2010), "Bent Coin Toss Probability," *Journal of the Korean Data and Information Science Society*, 21, 147–153. [2]