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

**František Bartoš[1*], Alexandra Sarafoglou[1], Henrik R. Godmann[1], Amir Sahrani[1], David Klein Leunk[2], Pierre Y. Gui[2], David Voss[2], Kaleem Ullah[2], Malte J. Zoubek[3], Franziska Nippold, Frederik Aust[1], Felipe F. Vieira[4], Chris-Gabriel Islam[5,6], Anton J. Zoubek[7], Sara Shabani[8], Jonas Petter[1], Ingeborg B. Roos[9], Adam Finnemann[1,10], Aaron B. Lob[2,11], Madlen F. Hoffstadt[1], Jason Nak[2], Jill de Ron[2], Koen Derks[12], Karoline Huth[2,13], Sjoerd Terpstra[14], Thomas Bastelica[15,16], Magda Matetovici[2,17], Vincent L. Ott[2], Andreea S. Zetea[2], Katharina Karnbach[2], Michelle C. Donzallaz[1], Arne John[2], Roy M. Moore[2], Franziska Assion[1,8], Riet van Bork[2], Theresa E. Leidinger[2], Xiaochang Zhao[2], Adrian Karami Motaghi[2], Ting Pan[1,9], Hannah Armstrong[2], Tianqi Peng[2], Mara Bialas[20], Joyce Y.-C. Pang[2], Bohan Fu[2], Shujun Yang[2], Xiaoyi Lin[2], Dana Sleiffer[2], Miklos Bognar[21], Balazs Aczel[22], and Eric-Jan Wagenmakers[1]**

[1]Department of Psychological Methods, University of Amsterdam,
[2]Department of Psychology, University of Amsterdam,
[3]Institute of Economic Law, University of Kassel,
[4]Faculty of Psychology and Educational Sciences, KU Leuven,
[5]Chairs of Statistics and Econometrics, Georg August University of Göttingen,
[6]Centre for Environmental Sciences, Hasselt University,
[7]Gutenberg School Wiesbaden,
[8]Department of Psychology, Justus Liebig University,
[9]Department of Adult Psychiatry, Amsterdam University Medical Centre,
[10]Centre for Urban Mental Health, University of Amsterdam,
[11]Cognitive and Behavioral Decision Research, University of Zurich,
[12]Center of Accounting, Auditing & Control, Nyenrode Business University,
[13]Department of Psychiatry, Amsterdam UMC, University of Amsterdam,
[14]Institute for Marine and Atmospheric Research Utrecht, Utrecht University,
[15]Centre Hospitalier Le Vinatier,
[16]INSERM, U1028; CNRS, UMR5292, Lyon Neuroscience Research Center, Psychiatric Disorders: from Resistance to Response Team,
[17]Research Institute of Child Development and Education, University of Amsterdam,
[18]Faculty of Behavioral and Movement Sciences, Vrije Universiteit Amsterdam,
[19]Department of Develomental Psychology, University of Amsterdam,
[20]Behavioural Science Institute, Radboud University,
[21]Doctoral School of Psychology, ELTE Eotvos Lorand University,
[22]Institute of Psychology, ELTE Eotvos Lorand University

*Correspondence concerning this article should be addressed to František Bartoš at f.bartos96@gmail.com.
For all but the last four positions, the authorship order aligns with the number of coin flips contributed.

## ABSTRACT

Many people have flipped coins but few have stopped to ponder the statistical and physical intricacies of the process. In a preregistered study 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—DHM estimated the probability of a same-side outcome to be about 51%. Our data lend strong support to this precise prediction: the coins landed on the same side more often than not, $\Pr(\text{same side}) = 0.508$, 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: $\Pr(\text{heads}) = 0.500$, 95% CI $[0.498, 0.502]$, $\text{BF}_{\text{heads-tails bias}} = 0.182$. Furthermore, this lack of heads-tails bias does not appear to vary across coins. Additional exploratory 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. Our data provide compelling statistical support for the DHM physics model of coin tossing.


## Introduction

Get attention w/ relevant "hook"

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 [1, 2, 3, 4, 5] and constituted a game of chance ('capita aut navia'–'heads or ships') already in Roman times (6, $\sim$ 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 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; the winner of the European Championship semi-final soccer match between Italy and the Soviet Union (an event which Italy went on to win) in 1968; which of two companies would be awarded a public project in Toronto in 2003; and to break the tie in local political elections in the Philippines in both 2004 and 2013.

Theoretical Models -- Basis of work

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 [3]. According to the standard model of coin flipping [7, 8, 9, 10, 11], 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; [13]) 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 non-zero 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) [13] 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.

Unknown/ Problems with existing data

Throughout history, several researchers have collected thousands of coin flips. In the 18$^{\text{th}}$ century, the famed naturalist Count de Buffon [14] collected 2,048 uninterrupted sequences of 'heads' in what is possibly the first statistical experiment ever conducted. In the 19$^{\text{th}}$ century, the statistician Karl Pearson [15] flipped a coin 24,000 times to obtain 12,012 tails. And in the 20$^{\text{th}}$ century, the mathematician John Kerrich [1] flipped a coin 10,000 times for a total of 5,067 heads while interned in Nazi-occupied Denmark. These experiments do not allow a test of the DHM model, however,

[1] Some even assert that a biased coin is a statistical unicorn—everyone talks about it but no one has actually encountered one [5]. Physics models support this assertion as long as the coin is not bent [12] or allowed to spin on the ground [3, 4].

Previous study

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 [16]: Larwood always started the flips heads-up, and Ku always tails-up. Unfortunately, the results (i.e., 10,231/20,000 heads by Larwood and 10,014/20,000 tails by Ku) do not provide compelling evidence for or against the DHM hypothesis.

Method preview

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 (and for individual differences in the extent of this bias). Furthermore, the data yield moderate evidence for the complete absence of a heads-tails bias. Exploratory analyses suggest the presence of a practice effect, as the degree of same-side bias decreases with the number of flips. The appendices demonstrate that the same conclusion obtains under a wide range of alternative analysis strategies.

Results preview

## Methods

### Data collection

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 hours coin-flipping (see e.g., [blinded for review] 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 denomination 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 (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 [blinded for review] for the data and video recordings.

### Audit

We randomly sampled and audited 90 sequences of 100 coin flips each. We verified the existence of the video recordings (with occasionally missing video recordings due to file corruption or recording equipment malfunction) and attempted to re-code the outcome of individual coin tosses from the video recordings. We encountered video recordings of varying quality and detail which made one-to-one matching of the original coded sequences and the re-coded audited sequences highly challenging. However, assessing the degree of same-side bias on the original vs. the audited sequences revealed that the original sequences contained a highly similar degree of same-side bias. As such, it seems implausible that the original sequences were affected by coding bias in favor of the same-side hypothesis.

## Statistical analysis

The preregistered and exploratory analyses outlined below are all Bayesian and use informed prior distributions. Appendix C provides a frequentist treatment. The Bayesian and frequentist results lead to the same qualitative conclusions.

[2] Including 2,700 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.

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