MAKING HEADS AND TAILS OF GENERATION LOSS: A TIMELESS TALE OF FOLK RANDOMNESS

Amy Renelle[1], Stephanie Budgett[1], and Egan Chernoff[2]
[1]The University of Auckland
[2]The University of Saskatchewan
amy.renelle@auckland.ac.nz

From the novel perspective of folk randomness, we document a journey through history that provides a lens for scrutinising the impact of changes to a well-known probability exercise—one involving a psychic professor undoubtedly identifying the fake sequence—and the implications of this for the learning opportunities it promotes. Evaluating iterations from the last 70 years, there is evidence of simplification that may have drawn this exercise away from its origins to the detriment of the task's original intentions. Importantly, this paper calls for statistics educators to rediscover the purpose of classroom tasks and retain the value we wish to impart—protecting the value in our statistical activities as they are morphed and mangled by time, translations between fields, educational focus, and underlying purpose.

INTRODUCTION

This novel study of and into folk randomness (explained below) evaluates a journey through history that provides a lens for scrutinising the impact of changes to a task and the implications of this for the learning opportunities it promotes. After introducing the task, tentatively called "The Coin Toss Comparison and the Psychic Professor," a brief history of the task from its roots in psychology to its now firm place in probability and statistics classrooms is presented.

In this paper, we coin the term "folk randomness." For the record, this is not a definition of randomness and is differentiated from other randomnesses such as "subjective randomness." For example, Batanero et al. (2016) captured the essence of subjective randomness by suggesting it is where a person's perception of randomness is dependent on the person's knowledge. Folk randomness, as we contend and we will demonstrate, emerges from our historical and critical examination of a task and is about the evolution of legends and the adaptation of the related lessons embedded within. In particular, we contend that *folk randomness* is the zeitgeist of randomness that emerges from the process of evaluating the purpose of a traditionally used probabilistic task when examining and considering historical changes over time.

THE COIN TOSS COMPARISON AND THE PSYCHIC PROFESSOR

This paper and, thus, the roots of folk randomness, concern the familiar activity that compares two sequences of coin flips, with one having been randomly generated and the other made up. In your own statistics classroom, you have likely heroically taken on the role of the psychic professor, predicting which sequence was made up by your students and which sequence was generated by coin flips. This is a well-established classroom tradition, often used to motivate students' interest in probability and statistics. The first rendition of this activity being used as a classroom exercise was detailed by Varga (1963, from Révész, 1990):

> *A class of school children is divided into two sections. In one of them, each child is given a coin which he throws two hundred times, recording the resulting head and tail sequence on a piece of paper. In the other section, the children do not receive coins but are told instead that they should try to write down a "random" head and tail sequence of length 200. Collecting all the slips of paper, he then tries to subdivide them into their original groups. Most of the time he succeeds quite well. His secret is that he has observed the following: In a randomly produced sequence of length two hundred, there are, say, head-runs of length seven. On the other hand, he has also observed that most of those children who had to write down an imaginary random sequence are usually afraid of writing down runs of longer than four. Hence, in order to find the slips coming from the coin tossing group, he simply selects the ones which contain runs longer than five.* (p. 95)

There are many iterations of this task, however. In our search, approximately 60 sources all present a rendition of The Coin Toss Comparison and The Psychic Professor Task. The changes made to the task throughout history, based on the results of our search, will now be presented.

In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina.* International Association for Statistical Education. iase-web.org

A BRIEF HISTORY

Although Varga presents the first classroom version of this activity, our historical examination begins back in 1949 when Reichenbach suggested that "[if] a person not trained in the theory of probability were asked to construct artificially a series of events that seems to him to be well-shuffled, there would not be enough *runs* in it" (italics as in original, p. 153). Likewise, also commenting on alternations of guesses, Solomon (1949) noted, "[it] has often been shown that organisms tend to avoid the repetition of responses in two-choice situations" (p. 322). It is posited that this could be the inspiration for the now well-established, yet morphed classroom exercise. The suggestions by Reichenbach (1949) and Solomon (1949) appear to draw upon three main areas of research from the early 20th century (see Tune, 1964): Fernberger's (1920, as referenced by Tune, 1964) repetition avoidance by human subjects comparing weights, the numerous experiments concerning repetition avoidance by rats in a maze (see Heathers, 1940, as referenced by Tune, 1964), and the Extra-Sensory Perception (ESP) experiments with a radio audience considered by Goodfellow (1938, as referenced by Tune, 1964). These studies all concluded that "organisms tend to avoid the repetition of responses" (Solomon, 1949, p. 322). It is this avoidance that may, based on the above, be seen in the psychic professor's classroom exercise—students typically create a pseudo-random sequence with runs that are too short and alternations that are too frequent.

Solomon's (1949) ESP experiment, designed to explore repetition avoidance rather than establish the existence of ESP, was conducted in a lecture with 192 psychology students. Holding a stack of five pennies in his hand, the professor would try to 'send' information to the students via ESP as to whether each penny was a head or a tail. The participants would then record their guesses. While this was an experiment with psychology students, rather than a statistics classroom demonstration to introduce probability concepts, there is certainly evidence of Solomon's experiment influencing today's rendition of the task. Subsequently, numerous experiments explored repetition avoidance. From the 1950's through to the late 1980's, these experiments primarily involved university psychology students. Although Varga's classroom exercise first demonstrated the use of this task with mathematics and statistics students, we have found that it took two decades for this task to be used as a classroom exercise with mathematics students. At the turn of the millennia, The Coin Toss Comparison and The Psychic Professor Task was no longer used as an activity with a wide range of participants and instead became solely a classroom activity with mathematics and statistics students (Figure 1 and Figure 2).

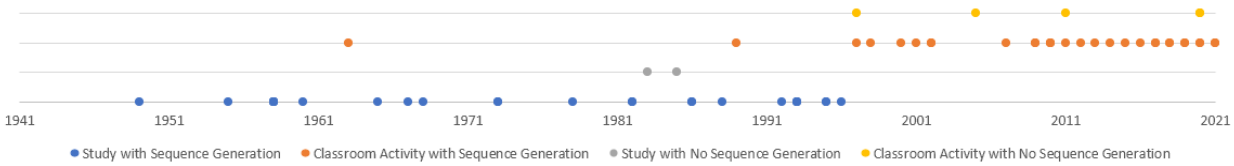

Figure 1. Example type (study or activity) and whether the task involved sequence generation

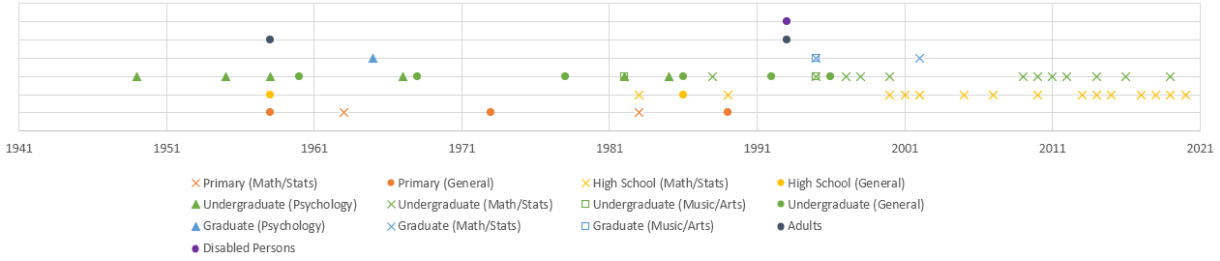

Figure 2. Participant background in The Coin Toss Comparison and The Psychic Professor Task from 1949 to 2021

Until 1983, all documented activities asked participants to generate a sequence. Since then, we found only six renditions providing participants with a sequence to consider. Similarly, while pre-2000 activities used both binary and non-binary contexts, and involved coins, cards, numbers, wheels, urns, and lights (Figure 3), after 1995 only binary contexts were used and, from 1997, the coin-toss context

was used exclusively. Worthy of note, post-2000 exercises often used nicely rounded numbers for the number of observations in the sequence considered, whereas pre-2000 there was an abundance of variation in the number of observations (Figure 4). Although sequences of length 100 and 200 have continually been used throughout the history of this task, pre-2000 activities also used other values (e.g., 6, 45, 64, 2520). It is likely that the use of nicely rounded numbers and the coin-based context increased in popularity due to the mathematics behind the ratio of heads and tails that make it possible for the professor to easily feign psychic abilities—such as the notion that students shy away from runs of more than length six in a sequence of length 200, so runs longer than this are a sign that the sequence is likely to be real rather than fake (Kouritzin et al., 2008; Renelle et al., 2021b; University of Arizona, 2011).

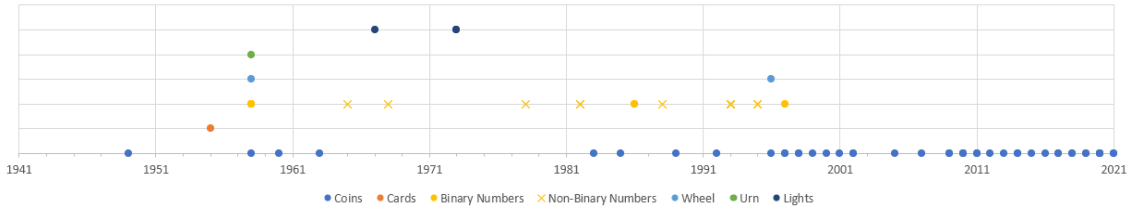

Figure 3. Use of different contexts in The Coin Toss Comparison and The Psychic Professor Task from 1949 to 2021

Figure 4. Number of observations used in The Coin Toss Comparison and The Psychic Professor Task from 1949 to 2021

Green's (1983, as referenced by Batanero & Sanchèz, 2005) example of this problem asked which sequence appeared "made up," whereas other non-generation-based tasks also worded the activity as which sequence was "produced by a fair coin" or "fake" (Figure 5). Particularly with lexically ambiguous terms such as those relating to randomness, the way the scenario is discussed and, therefore, the descriptor used, may impact how students carry out the activity (Renelle et al., 2021a). Typically, when describing the task, pre-2000 researchers used terms that corresponded to "sensing" randomness (such as those referring to ESP), "generating" randomness (i.e., with a fair coin in mind), or "behaving" randomly. Post-2000, "fake" and "invent" are more commonly used descriptors.

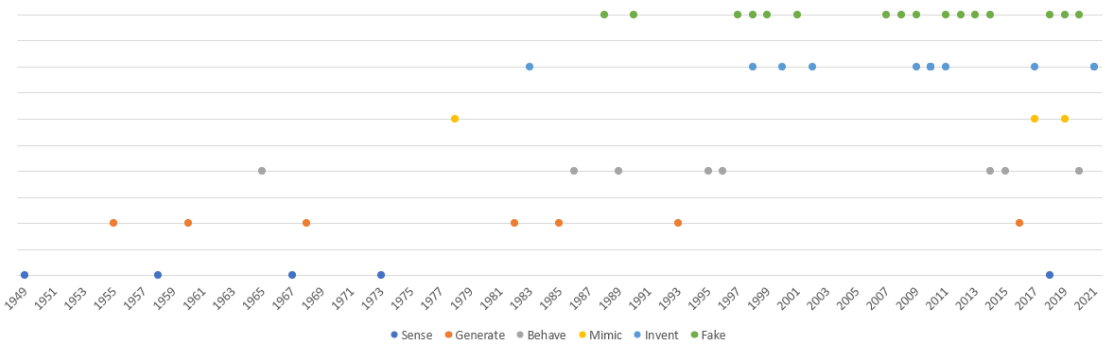

Figure 5. Descriptor use in The Coin Toss Comparison and The Psychic Professor Task

With such a rich history, a reflection of the adaptations of a task such as The Coin Toss Comparison and The Psychic Professor may offer valuable insight. Exploring the history of this task over the last 70 years, five key themes will be considered. There is evidence of a transformation of task intention across time: from mathematical tests to perceptions of randomness; from a generation task to a recognition task; from psychology experiments with university students as subjects to a statistics problem found in academic papers to a classroom opener for high school students starting probability. Descriptor terminology is also inconsistent between iterations of the task, particularly between fields, across time, and depending on the underlying focus of the task. There are, however, some parts of the task that seemed to stick. For example, a frequentist viewpoint seems to have been adopted at some point during this time, with 100 flips being the predominant sample size. There is also a strong preference for binary-based contexts, although this is not necessarily a fixed part of the task.

METHODS

As a novel approach to examining tasks, the process for conducting an analysis through the lens of folk randomness is still in development. Preliminary methods involved a thorough search of written examples found online. This included: books (n = 4), journal articles (n = 31, including 2 systemic reviews and 1 academic poster), media sources (n = 17, including news articles, blog posts, personal websites, and social media), and teaching resources (n = 8). An example of each kind of resource and the kind of data collected can be found in Table 1. A record, like this table, is kept of key characteristics to be considered, including the primary focus of content (long runs, alternations, biases, etc.), device implementation (dice versus coins, digital versus physical), sample size, and context (background). Comparison is then made between different groups and how they present the exercise, such as taking into account country variations and educator background (mathematics, statistics, education level, etc.). Future research will specify the method of folk randomness in more detail so the lens might be utilised by other researchers.

Table 1. Example of data collection for each resource type

Type	Sequence Generation	Exercise Type	Participant Background	Context	Number Observed	Descriptor Used	Reference
Book	Yes	Classroom	High School (Statistics)	Coins	100	Fake	Gelman & Nolan (2002)
Journal Article	Yes	Research Study	Undergraduate (Psychology)	Lights	288	Predict	Restle (1967)
Media Source	Yes	Classroom	High School (Statistics)	Coins	50	Behave	Lochel (2014)
Teaching Resource	Yes	Classroom	High School (Statistics)	Coins	50	Fake	New Zealand Ministry of Education (2018)

FOLK RANDOMNESS—A NEW PERSPECTIVE FOR STOCHASTICS EDUCATION

Having a sense of the history behind The Coin Toss Comparison and The Psychic Professor Task, we now discuss how we might draw meaning through use of our novel lens of folk randomness. As a lens with which to examine tasks and reflect on the changes particularly within statistics education, folk randomness utilises a historical examination of a task to uncover how the intentions of such a task have changed. Essentially, the lens of folk randomness allows a critique of the purpose and details of a task as it is currently implemented compared to previous and potential future iterations.

The Coin Toss Comparison and The Psychic Professor Task began as an activity used in psychology research to explore repetition avoidance, as we have shown. It was not until the turn of the millennia that the majority of authors publishing examples of this activity were from the field of

mathematics and statistics. Prior to then, we found only two such research groups in statistics education who conducted studies investigating this task; Green (1983, referenced by Batanero & Sanchèz, 2005) and Hill (1988). Green explored this activity as "part of his test on probability intuitions to 11 to 16-year-old children" (Batanero & Sanchèz, 2005, p. 245), finding that "…most of the students chose the nonrandom sequence" (p. 246) with reasoning that referenced the length of runs, the ratio of heads and tails, and patterns. This activity asked students to select the random sequence from two options, meaning this is not a sequence generation problem. Hill (1988) instead focused on the First Digit Phenomenon with undergraduate calculus students who were asked to write down a six-digit random number. While this activity was a sequence generation problem, the sequence produced was shorter than typical and the study focused on one aspect of sequence generation (that the first digit is more likely to be a small number). Along with Varga's classroom activity (1963, from Révész, 1990), these two studies, we have found, are the first to try the activity with mathematics and statistics students.

Of course, the overlap here is between the psychology of repetition avoidance and the heuristics and biases detailed by Tversky and Kahneman (1974). Following the publication of their landmark paper, a collection of research was published linking Tversky and Kahneman's heuristics and biases to renditions of The Coin Toss Comparison and The Psychic Professor Task. This is where the question appears to have changed from *"do humans avoid repetition?"* to *"is the representativeness heuristic affecting this person's judgement?"* The first instance of Tversky and Kahneman's work being referenced in relation to studies implementing The Coin Toss Comparison and The Psychic Professor Task appears to be Diener and Thompson (1985), who suggested that "[subjects] using a representativeness heuristic to classify series of heads and tails would attempt to determine directly whether each of the series was representative of a coin-tossing process" (p. 437). Their study invited undergraduate psychology students to consider whether each of the 50 pre-generated sequences were "… produced by tossing a fair coin …" (p. 439). Diener and Thompson found that, "If subjects relied on a representativeness heuristic to classify the series, it would be expected that the series most representative of a random process (those with the highest confidence ratings) would be classified more quickly than less representative series" (p. 445). Between 1985 and 2000, several researchers also reference Tversky and Kahneman in relation to The Coin Toss Comparison and The Psychic Professor Task, but none appeared to suggest that the representativeness heuristic should not be used when choosing which sequence was randomly generated or when generating a seemingly random sequence.

As the 2000's started, those in the field of mathematics and statistics appeared to decide that this task was suitable as a classroom activity, asking predominantly high school and undergraduate students to "invent" a "fake" random sequence of coin flips with 100 or 200 observations. In some instances, it seems that the aim of the activity was to demonstrate that humans are bad at creating randomness (e.g., Gamblers' Fallacy and Hot Hand Fallacy), that randomness may produce results different than expected, or to devise tests or rules for identifying whether a sequence is random or not. It is well established that humans are bad at "producing" randomness, and hence it is important to recognise disparities between short-run observations and theoretical probabilities. However, it seems that the purpose of this task sometimes is to demonstrate that the representativeness heuristic is an unhelpful bias for identifying or producing randomness. With either purpose in mind, many recent examples of this activity do not state this aim directly or expand upon it, failing to relate this idea to wider concepts and exemplify the consequences of these lessons.

Indeed, the lens of folk randomness has identified numerous aspects of this task that may appear malleable to educators looking for a fun activity with these changes potentially affecting students' learning moments. The aim of implementing the task could vary between focussing on heuristics and biases and the connection between simulated and theoretical probabilities. Given space limitations, some key task characteristics are considered very briefly in the next section.

DISCUSSION

Our historical narrative of this well-known task, presented through the novel lens of folk randomness, offers the opportunity for contemplation. As a valuable learning activity, what are the important characteristics to consider when planning to implement the task in the classroom? It is likely that the design of the task will depend on the intended learning outcomes for students. Does the physical random-generating device used in the task, for example coins, cards, lights, wheels, urns, die, adequately support the anticipated learning opportunities? How long should the generated sequences be? For

example, does a sequence with a given length invite students to adopt a "spot the longest run" approach to arrive at a deterministic conclusion regarding the "randomness" of the sequence? With technology now prevalent in statistics classrooms, what design principles would support a virtual random-generating activity? From a motivational perspective, are students more engaged when they, rather than the instructor, become the judge of the task? When students become recognisers of randomness rather than creators of randomness, how might this role-reversal impact learning? And how does the language used in the task align with the students' role (as recogniser or creator) and the aim of the activity? We invite those wishing to make use of this fun activity to ponder our reflections and to offer their own. As a well-known saying goes, the only constant is change. Alternatively stated, although the details, purpose, etc. of The Coin Toss Comparison and The Psychic Professor Task will always change over time, past, present, and future, that is, the task remains.

REFERENCES

Batanero, C., Chernoff, E. J., Engel, J., Lee, H. S., & Sánchez, E. (2016). *Research on teaching and learning probability. ICME-13 topical surveys*. Springer. https://doi.org/10.1007/978-3-319-31625-3_1

Batanero, C., & Sanchèz, E. (2005). What is the nature of high school students' conceptions and misconceptions about probability? In G. A. Jones (Ed.), *Exploring probability in school. Challenges for teaching and learning* (pp. 241–266). Springer. https://doi.org/10.1007/0-387-24530-8_11

Diener, D., & Thompson, W. B. (1985). Recognizing randomness. *The American Journal of Psychology, 98*(3), 433–447. https://www.jstor.org/stable/1422628

Gelman, A., & Nolan, D. (2002). *Teaching statistics: A bag of tricks*. Oxford University Press.

Hill, T. P. (1988). Random-number guessing and the first digit phenomenon. *Psychological Reports, 62*(3), 967–971. https://doi.org/10.2466/pr0.1988.62.3.967

Kouritzin, M. A., Newton, F., Orsten, S., & Wilson, D. C. (2008). On detecting fake coin flip sequences. *Institute of Mathematical Statistics Collections*, *4,* 107–122. https://doi.org/10.1214/074921708000000336

Lochel, B. (2014, December 5). *Class opener–day 64–can my students be random?* [Blog Post]. MathCoachBlog. https://mathcoachblog.com/2014/12/05/class-opener-day-64-can-my-students-be-random/

New Zealand Ministry of Education. (2018, September 10). *Activity: Fooling the teacher*. TKI: Te Kete Ipurangi. https://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Learning-programme-design/Year-13-programme-design/Level-8-statistics-example/Activity-Fooling-the-teacher

Reichenbach, H. (1949). *The theory of probability: An inquiry into the logical and mathematical foundations of the calculus of probability* (E. Hutton & M. Reichenbach, Trans.; 2nd ed.). University of California Press.

Renelle, A., Budgett, S., & Jones, R. (2021a). Defining randomness? In A. A. Bilgin & S. Budgett (Eds.), *Proceedings of the 10th Australian Conference on Teaching Statistics* (pp. 36–41). https://iase-web.org/documents/ANZCOTS/OZCOTS_2021_Proceedings.pdf?1637055230

Renelle, A., Budgett, S., & Jones, R. (2021b). New Zealand teachers' generation problem misconceptions. *Teaching Statistics*, *43*(2), 56–61. https://doi.org/10.1111/test.12248

Restle, F. (1967). Grammatical analysis of the prediction of binary events. *Journal of Verbal Learning and Verbal Behavior, 6*(1), 17–25. https://doi.org/10.1016/s0022-5371(67)80042-2

Révész, P. (1990). Regularities and irregularities in a random 0, 1 sequence. *Statistical Papers, 31*(1), 95–101. https://doi.org/10.1007/BF02924680

Solomon, R. L. (1949). A note on the alternation of guesses. *Journal of Experimental Psychology, 39*(3), 322–326. https://doi.org/10.1037/h0056986

Tune, G. S. (1964). Response preferences: A review of some relevant literature. *Psychological Bulletin, 61*(4), 286–302. https://doi.org/10.1037/h0048618

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

University of Arizona. (2011). *Real vs. fake data (for teachers)*. Institute for Mathematics and Education: G-Teams. http://ime.math.arizona.edu/g-teams/Profiles/DL/coinflips.pdf