One chip, two chip, red chip, blue chip: Re-examining the *equiprobability bias* as an expression of statistical expertise.

Research Question: Is the *equiprobability bias* (Lecoutre, 1992) really a cognitive bias, or do the results reported in Lecoutre (1992) represent an individual's level of statistical expertise?

Background and Purpose:

Whether you are a statistician or not, there is a good chance you encounter and interpret statistical information in your everyday life. From COVID-19 infection rates to fantasy football, we all are tasked constantly with not only interpreting statistical information in the day to day, but also making our different decisions based on them. However, despite being so prominent in our daily lives, many people still struggle to grasp the concepts of statistics, even after having completed a college level statistics course (Delmas, Garfield, Ooms, & Chance, 2007).

Why is statistics so difficult for people to understand? Tversky & Kahneman (1974) suggest that it is because people come hardwired with cognitive biases or heuristics. These biases, or heuristics, are simple, easy-to-apply rules that people use to make decisions, particularly when under a time constraint or when lacking enough information to make a more informed decision. Could a consequence of using heuristics be that it leads to systematic misinterpretations of statistical information? Lecoutre was one statistician of many that seemed to agree. In her 1992 paper she came to devise different types of probability tasks that she then presented to participants. The data she obtained led her to identify a new type of bias that she believed impacts our judgment of probability style questions: the *equiprobability bias* (Lecoutre, 1992). The equiprobability bias is a phenomenon in which people perceive the outcome of a random event to be equally likely simply because the event is random and even if that outcome is not necessarily the case. One of the most notable experiments which led her to conclude this was a simple type of token task. Imagine three poker chips in a box: two red, one white. If you close your eyes and draw two chips out of the box (at the same time), what combination of colors (*red red* or *red white*) do you think is most likely? Out of the answer choices*, equally likely, red and white and two red*, 44% of participants in Lecoutre's study said it would be equally likely, even though drawing a red and white is twice as likely as drawing two red.

Intrigued, a graduate student in the Sarnecka Lab tried to replicate the findings, however she came to observe different results. In her sample, fewer participants came to respond *equally likely* (17% vs. 44%) , more participants came to respond that they are *unable to answer* (25% vs. 4%) and more (albeit to a lesser extent) came to respond with the correct answer choice (15% vs. 6%). These results thus led her to believe that perhaps the phenomenon originally reported by Lecoutre (1992) might have in fact been partially mischaracterized. She hypothesized that students might instead just be more likely to use the bias when they have a weak background in statistics (i.e., are a novice), and less likely to use the bias when they have a strong background in statistics (i.e., are an expert), and Gebotys and Claxton-Oldfield might agree. In their 1989 paper they point out that errors in probability judgment might not necessarily be the result of using some type of heuristic, but instead the result of a person's weak knowledge base on probability. They also show how Tversky and Khanneman's (1972) work on heuristics is at least partly confounded with an individuals' level of probability expertise (Gebotys & Claxton-Oldfield, 1989)– could this be also true of the results reported in Lecoutre (1992)?

What does it even mean to be an expert, and how might the level of a person's expertise impact how they think about probability? A person is considered an expert in a domain when they have more knowledge in that given domain than their novice counterparts (Frederick & Libby, 1986; Gebotys & Claxton-Oldfield, 1989). In one study, Frederick and Libby (1986) demonstrated the effects of this knowledge gap by showing how novice and expert auditors come to systematically make different judgments on the effects of auditing errors as a level of their expertise. Gebotys and Claxton-Oldfield (1989) then came to show how a person's level of expertise also lessened their tendency to use decision strategies, such as the availability and representativeness heuristics, that lead to incorrect interpretations of probabilistic information (see Tversky & Kahneman, 1974). They also showed that just by improving a participants' knowledge of probability with a 15-minute learning session will make them less likely to rely on those aforementioned heuristics to make decisions on presented probabilistic information. Essentially, their intervention erased any novice/expert distinction that initially seemed to exist.

Novices' problem-solving and decision strategies also differ from those of experts. Experts generally use more selective search strategies when looking for information (Carmerer & Johnson, 1997). In the case of chess players, experts encode information about a game more efficiently than novices. This allows them to then identify the relevant information of a game more quickly and with subsequent ease (Chase & Simon, 1973). Additionally, because experts have such a large amount of domain-specific knowledge they can navigate to then identify the specific subset of relevant information they need to solve a task, they also generally end up using *less* information to arrive at a solution than their novice counterparts (Libby & Frederick, 1990; Shanteau, 1992).

Experts are also more likely to use concept-oriented solutions than novices (Hauff & Fogarty, 1996; Kozma & Russell, 1997). In their 1996 paper, Hauff and Fogarty found that successful statistics students were more likely to produce simple factual statements and adopt concept-oriented solutions to common statistics problems than unsuccessful statistics students. Similarly, Kozma and Russell (1997) found that when asked to create groupings of multimedia chemistry representations (e.g., graphs, equations, video clips), student experts created groupings based on conceptual relatedness while novices created the groupings based on surface-level features (e.g., visual features).

Given this background, could it be that Lecoutre's (1992) results reflect the use of a heuristic, or, could it instead be reflecting students that are developing a more nuanced, and correct, understanding of statistics? In an attempt to investigate this, we recruited undergraduate students at UCI and had them answer questions that were similar in style to the original Lecoutre token task. The students also completed a probability questionnaire (Gebotys & Claxton-Oldfield, 1989) that helped categorize them based on level of expertise (separate from might be indicated by the number of college level statistics courses taken, given that previous research suggests one might not be able to predict the other (Delmas, Garfield, Ooms, & Chance, 2007).

In the data collected, we found that as levels of expertise increased so did the prevalence of the equiprobability bias. We also found no relationship between expertise level and choosing the correct answer and that the more novice level participants were more likely to answer *red red* than *equally likely.*

This seems interesting, perhaps supporting Lecoutre's (1992) finds. But there were lots of problems with this data. Firstly, the data conformed as well to the null hypothesis as to the alternative (levels of expertise). The effect we found also did not seem to be that strong, and after a second look at the probability questionnaire, we found that a lot of the correct answer choices to the questions were actually *equally likely*. Essentially, the probability knowledge questionnaire was possibly not only playing into the *equiprobability bias*, but also into a specific knowledge level of the participant pool, which is counterintuitive given that we were using this questionnaire to specifically measure for expertise. Finally, even though we controlled for this last issue and found the same results, we still had participant probability scores that clustered around a single low value, indicating a high number of novices and a lower number of experts.

Analyzing this data further we noticed that the participants' expertise level seemed to track the correctness of the provided response. This "correctness" can be explained as follows: if we consider the true probability of each outcome, the probability implied by *equally likely* (p=.5) is closer to the probability of the correct answer *one red and one white* (p=.67), than the probability of *two red* (p=.33) is to the probability of the correct answer. Thus, the answer options can be ordered in terms of relative correctness such that *one red, one white* is completely correct, and *equally likely*, although incorrect, is more correct than *two red red*.

Most of the participants in our data were novices, and they either chose the answer choice *equally likely*, or the most incorrect answer choice –*two red*. Could it be that the increased prevalence of *equally likely* in Lecoutre's 1992 data and our own was simply a consequence of participants choosing an answer that was based on their level of expertise? In other words could it be that the *equiprobability bias* (Lecoutre, 1992) is not a cognitive bias in the way that Lecoutre (1992) discussed it to be?

We hypothesize that a participant's level of statistics expertise will predict the correctness of their answer choice in the token task, as opposed to their level of statistics expertise having no relationship with the correctness of their answer choice in the token task. If the equiprobability bias is a cognitive bias in the way that Lecoutre (1992) described, then there should be no relationship between participants' level of statistics expertise and the correctness of their answer choice on the token task. In other words, individual differences in level of expertise will not affect response accuracy on the task.The goal of the present project is thus to explore the relationship between statistical expertise and the tendency (or lack thereof) to use the equiprobability bias when interpreting statistical information. Specifically, the purpose of this proposed study is to repeat the experiment above with not only a wider subject pool of both novices and experts (through recruitment of undergraduates and also graduate students) but also with a new probability questionnaire using select questions from both the Gebotys and Claxton-Oldfield (1989) paper and the Gomez-torres et al. (2016) paper. Using the new data collected, we will then be able to more accurately see whether the results reported in Lecoutre (1992) seem to truly reflect a heuristic, or if they can more convincingly be explained by different levels of statistical expertise.

Method

Participants: I will recruit 111 UCI students through UCI's human subject research pool and classroom recruitment. This is the same number of participants I recruited in 2020 for an earlier version of this project where I observed a relationship between participants' statistical expertise and likelihood of using the equiprobability bias. Participants must be at least 18 years old. Participants will either receive a $5 Target gift card if they sign up to participate via classroom recruitment, or a ½ unit of credit though SONA for their participation.

Materials: There are 2 phases of the experiment. In the first phase, participants will view an image of 3 tokens and be asked to consider the likelihood of different events. Next, participants will complete a 14-item assessment of probability topics (Gebotys & Claxton-Oldfield, 1989; Gomez-torres et al, 2016). Samples of the stimuli are depicted in appendix: figure 1. The task will be administered online via REDCap hosted at UCI.

Procedure: Participants will sign up to participate via SONA or through their instructor. The study will be conducted online and asynchronously. At the start of the study, participants will be shown the study information sheet and asked to confirm that they meet the eligibility requirements described in the sheet. Upon confirmation, the study will begin. Before beginning the main task, participants will complete a brief demographic survey (e.g, age, gender, year in school). Then, the main task will begin. In the main task, participants will see the screen depicted in figure 1a and be asked to indicate which outcome they think is most likely to occur if two tokens are drawn at the same time: drawing two tokens of the same color, drawing two tokens of different colors, or the outcomes are equally likely. Then, participants will complete a basic probability questionnaire based on questions developed and/or used by Gebotys and Claxton-Oldfield (1989) and Gomez-torres et al (2016).

For participants who sign up through SONA, completion of the task and participation credit will be automatically tracked and assigned in SONA. For participants who sign up to participate through their instructor, participants will receive a confirmation code at the end of the study that they will email to a member of the study team. Upon receiving the confirmation code, a member of the study team will send the $5 Target gift card to the participant to their UCI email address.

Proposed Data Analysis: To assess whether participants' responses on the task reflect a cognitive bias or different levels of knowledge about probability, the data will be analyzed using multinomial logistic regression. This method will allow us to see how participants' level of statistics expertise, as measured by the 14-item assessment, predicts the probability of giving a certain response on the item in the main task.

Student responsibility: Undergraduate research assistants in the Sarnecka lab are expected to work a maximum of 10 hours per week in the lab: 2 hours per week are designated for lab meeting, and the remaining 8 hours will be allocated among the following activities:

- Drafting experimental stimuli
- Meeting regularly with my graduate student mentor and Professor Sarnecka
- Recruiting and monitoring student participation (SONA crediting & timeslots; organizing classroom recruitment)
- Data collection/ monitoring
- Assisting with data analysis
- Creating a poster for presentation at the UROP conference

Project Timeline:

Spring Quarter 2022:

- Turn in proposal to UROP office for review and approval (deadline: May 2, 2022)
- Update task with new probability assessment (estimated time: 1 week)

Fall Quarter 2022:

- Recruit participants through SONA and the classroom (pending approval from UROP)
- Collect data (target sample size: 111 participants, estimated time: 5 weeks)
- Analyze the data (estimated time: 2 weeks)
- Draft lab report for in-lab records (estimated time: 2 weeks)
- Create poster for presentation at UROP conference in Spring 2023 (estimated time: 2 weeks)

Itemized budget:

Item (Quantity)	**Cost**
1 poster (printing cost)	$80
Participant compensation	$555 ($5/participant, maximum of 111 participants)
Total cost	$635

IRB protocol number: N/A- protocol determined exempt by [Redacted](lead researcher) and [Redacted] (faculty sponsor) using the self-determination exempt tool. Completed tools and associated materials are on file with [Redacted].

References

Carmerer, C.F. & Johnson, E.J. (1997). The process-performance paradox in expert judgment. W.M. Goldstein, R.M. Hogarth, & L. Lopes (Eds.) In *Research on judgment and decision making: Current, connections, and controversies.* pp. 195-217. Cambridge University Press.

Chase, W.G. & Simon, A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55-81.

Delmas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6,* 28-58.

Frederick, D.M., & Libby, R. (1986). Expertise and auditors' judgments of conjunctive events. *Journal of Accounting Research, 24*, 270-290. doi: 10.2307/2491134

Gebotys, R.J. & Claxton-Oldfield, S.P. (1989). Errors in the quantification of uncertainty: A product of heuristics or minimal probability knowledge base? *Applied Cognitive Psychology, 3*, 157-170. doi: 10.1002/acp.2350030206

Gómez-torres, Emilse, et al. "Developing a Questionnaire to Assess the Probability Content Knowledge of Prospective Primary School Teachers." *STATISTICS EDUCATION RESEARCH JOURNAL*, vol. 15, no. 2, 2016, pp. 197–215., doi:10.52041/serj.v15i2.248.

Hauff, H.M. & Fogarty, G.J. (1996). Analysing problem solving behaviour of successful and unsuccessful statistics students. I*nstructional Science, 24*, 397-409. doi: 10.1007/BF00125577

Kozma, R.B. & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *Journal of Research in Science Teaching, 34*, 949-968. doi: 10.1002/(SICI)1098-2736(199711)34:9<949::AID-TEA7>3.0.CO;2-U

Lecoutre, M.P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics, 23*, 557-568. doi: 10.1007/BF00540060

Lee, M.D., & Sarnecka, B.W. (2010). A model of knower-level behavior in number concept development. *Cognitive Science, 34,* 51-67. doi: 10.1111/j.1551-6709.2009.01063.x

Levenberg, S.B. (1975). Professional training, psychodiagnostic skill, and kinetic family drawing. Journal of Personality Assessment, 39, 389-393. doi: 10.1207/s15327752jpa3904_11

Libby, R. & Frederick, D.M. (1990). Experience and the ability to explain audit findings. *Journal of Accounting Research, 28*, 348-367. doi: 10.2307/2491154

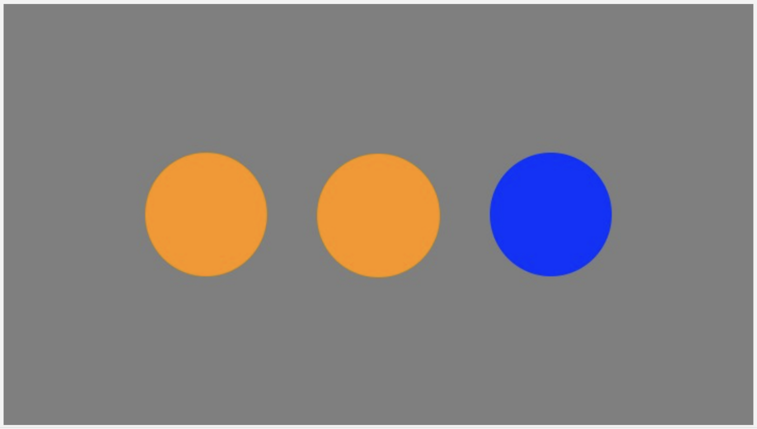
Sarnecka, B.W., & Lee, M.D. (2009). Levels of number knowledge in early childhood. *J Exp Child Psychol., 103*, 325-337. doi: 10.1016/j.jecp.2009.02.007

Shanteau, J. (1992). How much information does an expert use? Is it relevant? Acta Psychologica, 81, 75-86. doi: 10.1016/0001-6918(92)90012-3

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Appendix

In this box, there are three tokens; two are orange, and one is blue. I am going to draw two tokens. Do you think:



- ◯ The chance of obtaining one orange and one blue token is equal to the chance of obtaining two orange tokens
- ◯ There is more chance of obtaining one orange and one blue token
- ◯ There is more chance of obtaining two orange tokens
- ◯ Impossible to answer

reset

Why?

Expand

a.

Consider a coin, symmetric and delicately labelled 'head' on one side and 'tail' on the other. Toss the coin 10 times and record the number of heads. Toss the coin 100 times and record the number of heads. In which case would you expect the proportion of heads to be closer to .5?

- ◯ 10 toss case
- ◯ 100 toss case
- ◯ Both equally likely
- ◯ Don't know

reset

Toss a symmetric, delicately labelled coin (labelled 'head' on one side and 'tail' on the other) five times. Two possible sequences are given below, where "H" denotes "heads" and "T" denotes "tails":
(1) H H H H H
(2) H T T H T

Is:

- ◯ Outcome 1 most plausible?
- ◯ Outcome 2 most plausible?
- ◯ Both outcomes are equally plausible?
- ◯ Don't know

reset

Consider two urns containing red and black balls, identical except for color.
Urn 1: 80 red, 20 black
Urn 2: 10 red, 90 black

Select a ball at random from urn 1. What is the probability of obtaining a red ball?

b.

Figure 1. a.) Picture from phase 1 of the experiment. In this phase, participants see 3 tokens depicted on the screen. They will be asked to consider the likelihood of different outcomes and click on the response they think is most likely. They also have the option to explain their choice. b.) A sample of questions from the probability assessment that participants will complete in phase 2.