Educational Tools for Probabilistic Machine Learning Curriculum in Schools

Josephine Rey
Computer Laboratory
University of Cambridge
jmr239@cam.ac.uk

Alan F. Blackwell
Computer Laboratory
University of Cambridge
afb21@cam.ac.uk

Xinyue Li
Cambridge University
Press & Assessment
University of Cambridge
xinyue.li@cambridge.org

Gemma Penson
gpensonite@gmail.com

Hong Ge
Engineering Department
University of Cambridge
hg344@cam.ac.uk

Helen Arnold
Independent Tutor
helenlindaarnold@gmail.com

Abstract

As Bayesian approaches to probability and statistics become more widespread foundations of machine learning, there is interest in introducing basic principles of probabilistic modelling at secondary school level. This paper presents a series of educational experiments with simple probabilistic modelling tools based on probabilistic programming languages.

1. Introduction

This paper reports on progress within a long-term project, following earlier reports at the Psychology of Programming Interest Group (PPIG). The overall agenda is the use of probabilistic programming languages (PPLs) to enhance education in probability and statistics, specifically when introducing concepts of probabilistic modelling into school and university curricula.

This educational agenda was one focus of a paper at PPIG 2019 that introduced the research field of usability of PPLs (Blackwell et al., 2019), in a multi-authored paper including many of the contemporary leaders in PPL development and research. Among other contributions, the 2019 paper suggested that the field may benefit from a 'furthest-first' strategy (starting design work with those who are most excluded), in this case by undertaking initial scoping research with school students in remote and disadvantaged communities, in particular on the African continent. Early results reported on experiments with visualising Bayesian probability in the Kalahari (Blackwell, Bidwell, et al., 2021), and use of causal models by schoolchildren in Nigeria to estimate the risk of Covid-19 infection on the basis of observations (Attahiru, Maudslay, & Blackwell, 2022).

These educational experiments have been planned in collaboration with an international mathematics curriculum research team based at Cambridge University Press & Assessment (CUP&A), who have developed a framework of mathematical concepts and learning objectives encompassing a comprehensive range of probability and statistics content across both primary and secondary school curricula (Cambridge Mathematics, n.d.). In addition to their ongoing work, the team has recently begun investigating how to support skills and practices in the field of digital technology-enhanced education, including AI (Li & Zaki, 2024). One of the goals for this team is to anticipate and document future developments in this curriculum area that will enhance competencies and advanced learning in this evolving field, much of which is dependent on fundamental principles of probability and statistics that are not yet routinely included in school curricula (Slesinski & Fadel, 2024; Hoegh, 2020).

In this paper, we present two further projects advancing our investigation: one evaluating an interactive visualisation of Causal Bayesian Networks in a UK classroom context and the other extending the Scratch language with PPL functionality to investigate how classroom use of interactive statistical modeling tools in South Africa might support curriculum priorities in that country.

2. Previous Work

As described in the introduction, work that we have previously presented at PPIG explored foundational concepts in Bayesian probability through a 'furthest-first' agenda to engage communities historically excluded in curriculum research. This programme of work is in contrast to other initiatives that have introduced PPLs in university-level teaching, typically at an advanced level, including Oxford, Harvard and Columbia [1].

2.1. Conditional probability in the Kalahari

The first of these explorations investigated ways of representing and thinking about probability in relation to the context and needs of the Ju|'hoansi people living near Tsumkwe, Namibia (Bidwell et al., 2022). As hunter-gatherers, a strong ability to reason about likelihood from observed data enables survival and success. This work explored interactive visualisations of conditional probability, using physical spinners made from cardboard and paperclips to carry out simple Monte Carlo simulations that quantitatively explored the causal relationship between random variables (Blackwell, Bidwell, et al., 2021). For example, the chance of finding water under different temporal and situational scenarios (i.e. '*after rain*', '*within a tree*' and '*within a tree given that it may be home to a snake*') places the foundations of Bayesian ideas within indigenous knowledge practices and elevates the importance of making AI accountable to diverse knowledge practices (Bidwell et al., 2022).

2.2. Causal reasoning during a pandemic

Following interruption of the Kalahari fieldwork by the Covid-19 pandemic, we created a simple Javascript emulation of the cardboard spinner, allowing interactive Monte Carlo simulations to be explored remotely with our field research collaborator and translator on the screen of his Android phone (Blackwell, Bidwell, et al., 2021). Simulated outcome frequencies were tallied in an interactive webpage, with the proportion of different outcomes for each variable rendered as a pie chart whose sector sizes could be compared to the relative sizes of the spinner sectors as a demonstration of long-run probabilities.

These visualisations were used as the starting point for a classroom experiment in a school in Nigeria, where a lesson plan asked children to quantify their relative risks of being infected with Covid, as informed by observations they might make in a local market (Attahiru et al., 2022). Likelihoods of different outcomes for each random variable were again visualised as different-sized sectors in a pie chart. Causal relations between random variables were visualised as links between the pie charts, showing how the different outcome likelihoods of an unknown variable might be updated on the basis of observations of other variables that it is conditioned on.

Using these visualisations, a workshop with eight students was carried out remotely, using a web-based lesson plan trialled by a teacher known to the researcher in Nigeria. The static visualisations of likelihoods were not successful in this case, in large part because the lesson plans were neither clearly related to students' personal experiences of risk and likelihood nor to the standard curriculum in probability and statistics as taught at that level in Nigeria.

3. Interactive Visualisation of Causal Bayesian Networks

As an improvement over the static visualisation concept evaluated in (Attahiru et al., 2022), we created a dynamic version of the same visualisation, in which the circular nodes of a Bayesian network are again replaced by pie charts whose sector sizes correspond to the relative likelihoods of different outcomes for a categorical random variable. Figure 1 shows an overview of this system's operation.

The interactive network graphs allow students to create and link nodes, specify frequencies of known values for exogenous variables (observed data external to the model), and observe expected outcome distributions for the endogenous variables (values inferred by the model) that are specified as child nodes in the graphical model. The graph is constructed by interactively creating and linking nodes, with

[1] e.g. Oxford's *Bayesian Statistical Probabilistic Programming*, Harvard's *Probabilistic Programming and Artificial Intelligence*, Columbia's *Applied Statistics III Nonparametric Theory in Machine Learning*

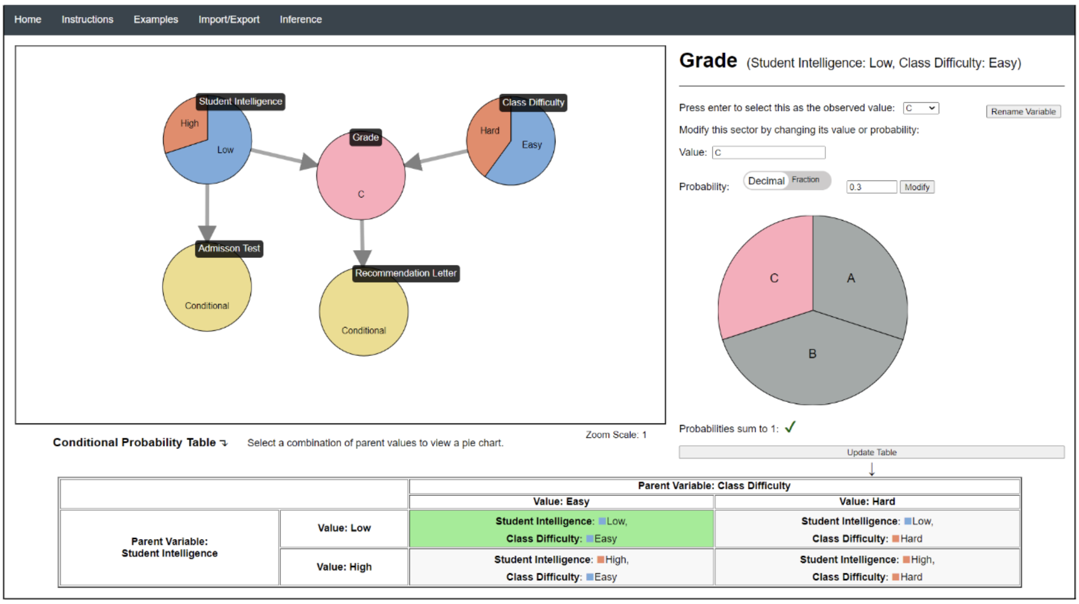

Figure 1 – Overview of the interactive CBN visualisation created by Penson, showing a Causal Bayesian Network with relative likelihood of values indicated by sizes of pie chart segments within each node. The bottom part of the display shows a corresponding conditional probability table.

class likelihoods for exogenous variables entered via an interactive dialog, and model updates rendered as modified pie charts for the endogenous variables.

3.1. Implementation

With the visualisation implemented in d3.js, the original plan had been to support more complex or data-intensive models via back-end execution in the Turing PPL (Ge, Xu, & Ghahramani, 2018). Early experiments also used the Julia implementation of BUGS, which shares Julia components with Turing.jl (Xianda Sun & Ge, 2024). However, performance issues with that server connection, and the relative simplicity of the teaching scenarios, meant that the classroom deployment of the system could be achieved with in-browser execution, with the graphical model compiled to the TypeScript PPL BayesJS[2] (re-compiled using Browserify). An example of the BayesJS node syntax corresponding to one of our teaching examples is shown in Figure 2.

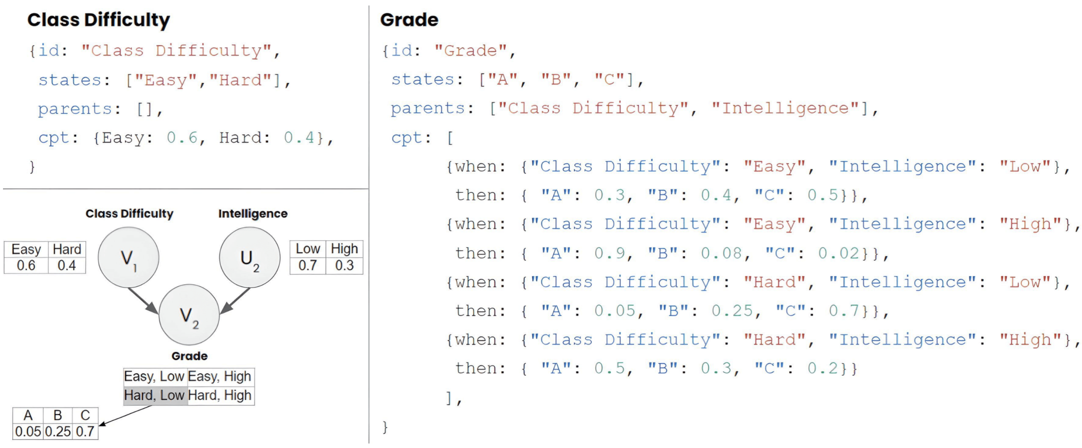

Figure 2 – BayesJS node syntax corresponding to part of the university applications model

[2] github.com/bayesjs/bayesjs

3.2. Evaluation

Where Attahiru's project (Attahiru et al., 2022) had explored causal reasoning in relation to experiences that were expected to be meaningful to schoolchildren in Nigeria, we evaluated this interactive visualisation in a Western context, with secondary school students studying the UK sixth-form curriculum in Further Mathematics. In this context, a cultural priority is how students will achieve the necessary grades for admission to university. We, therefore, used a teaching example that focused on causal factors in university admission, based on the Western approach to teaching Bayesian probability advocated by Pearl (Pearl, 1995), and replicating teaching examples that had previously been used for school-based research in other Western settings by (Lecoutre, 1992) and (Gordon, Henzinger, Nori, & Rajamani, 2014).

Classroom evaluation of our tool focused on the following research questions:

1. Do students employ Bayesian thinking in their decision-making? (Pre-intervention)
2. Can Bayesian thinking be induced through interaction with an interface? (Intervention)
3. Is the system's usability sufficient to support students completing Bayesian tasks? (Usability)

An in-school trial was conducted with 20 students at Hill's Road Sixth Form College in Cambridge. This took place during a timetabled class in Further Maths, supervised by a classroom teacher. The study was approved by the ethics committee of the Cambridge Department of Computer Science and Technology.

The research questions were investigated through six tasks:

- Tasks 1, 4, 5, and 6 directly address the research questions:
  - Task 1 (Pre-intervention): Students answer questions about BT without using the interface. (RQ1)
  - Task 4 and 5 (Intervention): Students build a Causal Bayesian Network (CBN) using either the interface or on paper (within-subjects). (RQ2)
  - Task 6 (Usability): Students complete self-directed tasks with the interface, reporting difficulty and completion level. (RQ3)
- Tasks 2 and 3 (Intervention) These provide foundational knowledge for later tasks:
  - Task 2: Lecture on probability and probability trees.
  - Task 3: Introduction to CBNs using examples.

The order of completing tasks with and without the interface is switched between groups to avoid bias. Overall, this evaluation aims to assess if the educational tool can effectively teach Bayesian thinking through interaction and if the interface itself is usable for students.

Using Likert scale measures in a post-intervention survey, students reported improved understanding of Bayes theorem, conditional and marginal probabilities, and prior and posterior likelihoods ($p < 0.05$). In comparing usability between the two presentation conditions, they reported that hand-drawn CBNs were easier and faster to create (Welch's t test, $p < 0.05$), while those created using the interactive editor were easier to modify and explore.

4. Teaching Bayesian probability in a South African context

This section reports a preparatory study exploring the potential of digital tools to introduce Bayesian concepts in probability education in South Africa. The first author conducted interviews with educators, curriculum designers, and NGO leaders to understand the challenges of designing future probability curricula and how the current South African Curriculum Assessment Policy Statements (CAPS) support digital tools and their use in classrooms (Department of Basic Education, 2011).

The study identified several challenges in South Africa's education system. The shift from a student's mother tongue to English in the Intermediate Phase (ages 9-13) creates difficulties for learners when acquiring subject-specific knowledge. Additionally, the CAPS curriculum is outcomes-driven and rigid in structure, with its delivery frequently affected by structural and socioeconomic inequality between schools (Spies, 2022).

Regarding digital tools, teachers lack the resources and training necessary to integrate technology effectively into their classrooms, and CAPS does not provide guidance for employing such tools. When considering future curricula, the importance of addressing infrastructural limitations and insufficient teacher training for digital tools becomes clear. The findings emphasised that curricula should celebrate African perspectives and integrate indigenous knowledge systems. However, they also acknowledge the need for a balance between including these local contexts and ensuring students are prepared for a globalised world.

The study also explored the challenges of teaching probability and mathematics in South Africa. Probability studies fall under the mathematics curriculum's 'Data Handling' focus area, but feature few real-world notions of likelihood beyond games of chance and simple data collection/analysis. Findings suggest that games from African cultures could be a valuable resource for teaching probability concepts instead of those related to suits and cards.

In designing digital tools to support future curricula, informants stressed the importance of accessibility for learners with varying digital literacy levels. Teacher training and support are crucial for the successful implementation of these tools. The study also highlights the need for the tools to function offline, considering limitations like power outages that are common in South Africa. Finally, user testing in classrooms is essential to evaluate and improve the effectiveness of these digital tools.

Overall, the findings highlighted the need for culturally-relevant curricula that integrate African knowledge systems alongside globally recognised educational standards. Inclusive digital tools are necessary, but researchers must address infrastructural challenges and ensure these tools can support diverse learners. The study emphasises the importance of ongoing collaboration between researchers, educators, and policymakers. To this end, South African educators are committed to improving education and building capacity for future skills, with efforts underway to integrate technology, social-emotional learning, and indigenous knowledge into the curriculum.

5. The ScratchTuring hybrid PPL

Previous experiments in this programme of work, as reported above, have achieved web-deployable prototypes by using visualisation libraries such as d3.js and implementing basic programmability with additional semantic elements such as the facility to link pie charts together as nodes of a graph. In these earlier systems, more sophisticated probabilistic modelling has been achieved either through back-end execution using a general-purpose PPL such as Turing or JuliaBUGS or in-browser execution via compilation to Javascript as in BayesJS.

In order to explore the potential for web-based educational PPLs in the South African context, we used a more powerful formalism for visual computation: the well-known Scratch language and environment originally created for school-level introductory programming classes (Resnick et al., 2009). Scratch already has sophisticated editing, interaction, and visualisation capabilities, and is deployed in a fully browser-based version with extension capabilities that enabled the extensions described below.

5.1. Turing interface

The ScratchTuring hybrid introduces new Scratch blocks that invoke the PPL functionality of the Turing.jl language, cross-compiled into Turing scripts that are executed on a (local or remote) server, which can then be queried from Scratch to visualise the probabilistic model. We created Scratch-syntax wrappers for a subset of the Turing language so that Turing development tasks, such as creating and conditioning models on new data, can be done within the Scratch editor. This functionality is delivered through a constrained set of blocks specifically to support lesson plans in probabilistic modelling, as seen in Figure 3.

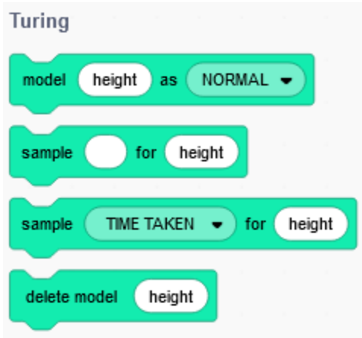

Figure 3 – Scratch blocks used to interact with model distributions in Turing

5.2. Geographical data

We wanted to address the policy demand in South Africa for educational tools that relate to recent advances in machine learning while also considering the geographical and cultural context within which South African students engage with such advances. We therefore created new Scratch capabilities relating to these concerns, in contrast to the original development and evaluation of Scratch that focused on the Western priorities of children's engagement with computer games and digital media.

One lesson plan was inspired by previous work that had provided schoolchildren in Ethiopia with programmable access (via simplified Python libraries) to satellite imagery from Google Earth Engine geotagged with the what3words API (Longdon, Gabrys, & Blackwell, 2024). We created a Scratch extension that loads a satellite image from any specified coordinates as a 'backdrop'. This image can then be used in conventional Scratch code, as seen in Figure 4 where the Scratch character is making a random walk, sampling colours from an image of the South African coast.

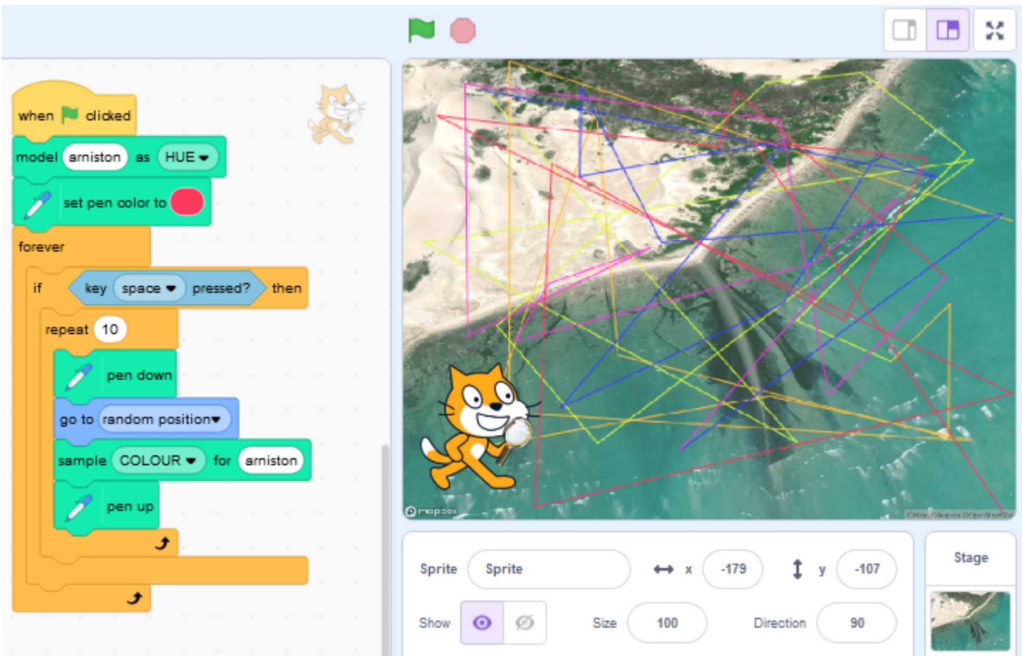

Figure 4 – The ScratchTuring interface used to create a program that collects samples from a satellite image

5.3. Probability model dashboard

The core extension that allows ScratchTuring to be used as an interactive tool for exploration of probabilistic models is a model *dashboard*, implemented as a new tab in the browser-based Scratch client. The operation of this dashboard can be seen in Figure 5, which shows a teaching application where each Scratch sprite represents an elephant in the Kruger National Park, South Africa's largest nature reserve. The elephant sprite reports its attributes as sample observations, updating a statistical model maintained in Turing. The ScratchTuring dashboard can then be used to visualise a probability density function reported by that model, showing the prior distribution before the observation, followed by (as in Figure 5) the posterior distribution. Using these facilities, a teacher projecting the dashboard can deliver exploratory interactive lessons, and students may also experiment with Scratch to create larger or more complex data science projects and simulations.

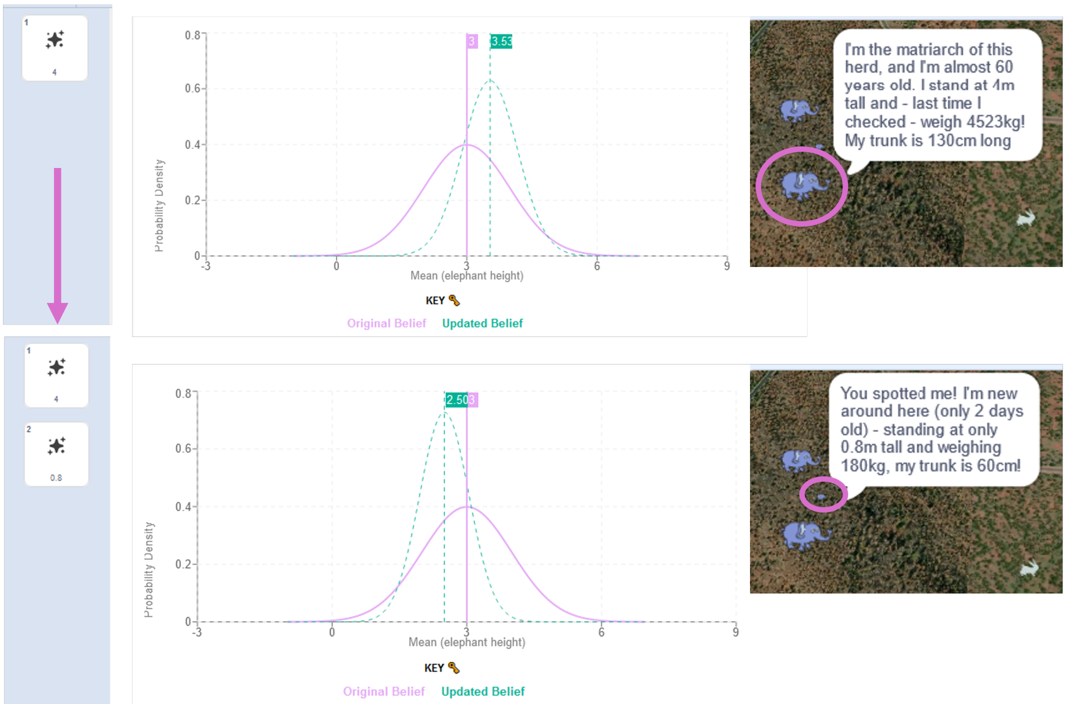

Figure 5 – The ScratchTuring model dashboard showing prior and posterior distributions based on reported elephant heights. After observing a baby elephant, the posterior distribution (green dotted line) reflects an updated belief that elephants can sometimes be smaller.

5.4. Visualising image samples as a hue distribution

As an experiment in relating local understanding of satellite image data to simple probabilistic methods in computer vision and machine learning, we implemented a model type and visualisation that renders probability distribution for colours sampled from a hue spectrum, as seen in Figure 6. The hue sampling Scratch block calculates a local average RGB from the background of a sprite's location and maps this into hue space (with saturation and brightness collapsed). This allows a distribution of hues to be calculated across samples from a satellite, as seen in figure 6. In the figure, it can be seen that hues are drawn from two different distributions, one corresponding to portions of the satellite image containing the ocean, and one corresponding to the colour of the beach. Although rendering a hue spectrum as the x-axis of a histogram is intuitively appealing, it should be noted that not all HSB colour values are easily perceived as being similar to the same hue value with 100% saturation and brightness. To help students appreciate this mapping, the histogram bars are rendered using the average of the saturation and brightness values observed, rather than the bright colours shown on the axis. The actual colours observed in the samples collected for a particular hue also pop up as a set of patches when the user hovers over its corresponding bar.

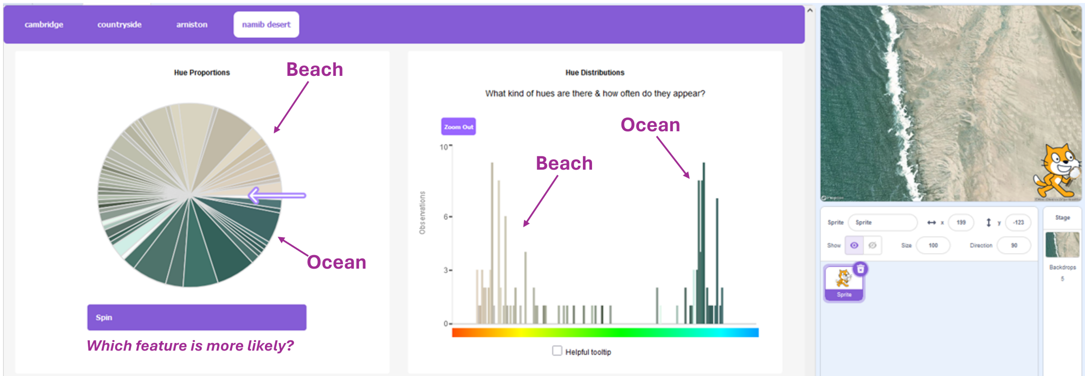

Figure 6 – The ScratchTuring model dashboard showing the distribution of hues sampled from a satellite image. The samples are collected using behaviour added to the Scratch character, shown here carrying a magnifying glass, whose lens is positioned over the part of the image from which the sample will be taken.

5.5. Lesson evaluation

At the time of writing, the ScratchTuring prototype has been presented to UK students at a specialist mathematics school in Cambridge. Students from the school are planning to travel to Kenya in the summer of 2024 for a programme of activities planned to include volunteer teaching at a Kenyan primary school. There may be an opportunity for students to use some aspect of this project during their visit.

The ScratchTuring prototype has also been demonstrated to several of the South African teachers who participated in the study described in Section 4. Additionally, a lesson plan incorporating this tool was recently introduced to Grade 8 students (ages 13-14) in Langa, a township in Cape Town, during a mathematics class at a specialist maths and science secondary school.

6. Discussion

We have described two interactive prototypes, continuing a series of educational experiments studying the potential use of probabilistic programming languages for teaching Bayesian probability and statistics at school level. Both systems are intended for web deployment in classrooms and present an interactive graphical front end, with a more conventional PPL used to construct and manipulate a Bayesian model.

We are especially interested in the potential for new developments in international school curricula to be initiated in non-Western contexts, including countries in Africa. Although the authors are based in the UK and often work with local schools as required by the practicalities of student research projects, we are focused on curriculum ideas that can be informed by local understanding and indigenous knowledge traditions from other parts of the world (Blackwell, 2021; Blackwell, Damena, & Tegegne, 2021). Since the methods of Bayesian probability underpinning recent advances in machine learning already represent a change in emphasis from the established conventions of frequentist statistics embedded in today's Western curricula (Slesinski & Fadel, 2024), this seems an ideal opportunity to consider the opportunities and implications arising less WEIRD (Western, Educated, Industrialised, Rich, Developed) ways of thinking (Henrich, Heine, & Norenzayan, 2010; Escobar, 2018).

At the 2022 PPIG workshop, Zainab Attahiru presented lesson plans that visualised conditional probability in a form intended to be accessible to students in Nigeria, allowing them to reason quantitatively about risks in their own lives (Attahiru et al., 2022). Gemma Penson's project, as reported here, has implemented an interactive version of the same visualisation, demonstrating that it can be used by UK students to reason in a probabilistic way about their school ambitions.

The series of educational experiments described in this paper aligns with key learning sciences theories such as project-based learning (engaging students in real-world challenges), situated learning (embedding education within its natural context), and simulation-based learning (using interactive, real-life scenarios). These perspectives afford the potential to extend the findings beyond local settings to international and global contexts.

The ScratchTuring prototype that we have introduced builds on these experiments to create a fully featured visual programming environment, with facilities supporting direct modelling of problem domains relevant to African learners. Preliminary teacher evaluation suggests that ScratchTuring is sufficiently robust and usable for deployment in classrooms, and we expect to be able to report on those deployments at the PPIG workshop.

7. Acknowledgements

We are grateful to the many teachers and students who have advised us and participated in the experimental lessons reported here. Josephine Rey's research has been funded by the Margaret & Patrick Flanagan Trust. Continued research enhancing the usability and educational applications of the Turing language is funded by the Alan Turing Institute. Generative AI declaration: The preparation of this manuscript employed multiple automated language-processing tools, including functions for spelling and grammar correction, summarisation, paraphrasing and predictive text.

8. References

Attahiru, Z., Maudslay, R. H., & Blackwell, A. F. (2022). Interactive bayesian probability for learning in diverse populations. In *Ppig* (pp. 77–87).

Bidwell, N. J., Arnold, H., Blackwell, A. F., Nqeisji, C., Kunta, K., & Ujakpa, M. (2022). Ai design and everyday logics in the kalahari. In *The routledge companion to media anthropology.* Routledge. Retrieved from `https://www.routledgehandbooks.com/doi/10.4324/9781003175605-54` (Accessed on: 14 Dec 2023)

Blackwell, A. F. (2021). Ethnographic artificial intelligence. *Interdisciplinary Science Reviews*, *46*(1-2), 198–211.

Blackwell, A. F., Bidwell, N. J., Arnold, H. L., Nqeisji, C., Kunta, K., & Ujakpa, M. M. (2021). Visualising bayesian probability in the kalahari. In *Ppig.*

Blackwell, A. F., Church, L., Erwig, M., Geddes, J., Gordon, A., Maria, I. G., ... others (2019). Usability of probabilistic programming languages. In *Ppig* (pp. 53–68).

Blackwell, A. F., Damena, A., & Tegegne, T. (2021). Inventing artificial intelligence in ethiopia. *Interdisciplinary Science Reviews*, *46*(3), 363–385.

Cambridge Mathematics. (n.d.). *Randomness and probability research summary.* (unpublished)

Department of Basic Education. (2011). *Curriculum and assessment policy statement: Mathematics grades 1-3.*

Escobar, A. (2018). *Designs for the pluriverse: Radical interdependence, autonomy, and the making of worlds*. Duke University Press.

Ge, H., Xu, K., & Ghahramani, Z. (2018). Turing: a language for flexible probabilistic inference. In *International conference on artificial intelligence and statistics* (pp. 1682–1690).

Gordon, A. D., Henzinger, T. A., Nori, A. V., & Rajamani, S. K. (2014). Probabilistic programming. In *Future of software engineering proceedings* (pp. 167–181).

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and brain sciences*, *33*(2-3), 61–83.

Hoegh, A. (2020, September). Why Bayesian Ideas Should Be Introduced in the Statistics Curricula and How to Do So. *Journal of Statistics Education*, *28*(3), 222–228. Retrieved 2023-11-06, from `https://doi.org/10.1080/10691898.2020.1841591` (Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10691898.2020.1841591) doi: 10.1080/10691898.2020.1841591

Lecoutre, M.-P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational studies in mathematics*, *23*(6), 557–568.

Li, X., & Zaki, R. (2024). Harnessing the power of digital resources in mathematics education: The potential of augmented reality and artificial intelligence. In S. Papadakis (Ed.), *Iot, ai, and ict*

for educational applications: Technologies to enable education for all (pp. 191–223). Cham: Springer Nature Switzerland.
Longdon, J., Gabrys, J., & Blackwell, A. F. (2024). Taking data science into the forest. *Interdisciplinary Science Reviews*, *49*(1), 82-103. doi: 10.1177/03080188241230415
Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, *82*(4), 669–688. Retrieved 2023-12-19, from `http://www.jstor.org/stable/2337329`
Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., ... others (2009). Scratch: programming for all. *Communications of the ACM*, *52*(11), 60–67.
Slesinski, J., & Fadel, C. (2024). *What "mathematics of ai" should be taught in schools?* (Tech. Rep.). Center for Curriculum Redesign.
Spies, A. (2022). A culture of exclusion: Re-configuring inclusive education in south africa. *International Human Rights Law Review*. Retrieved from `https://api.semanticscholar.org/CorpusID:253231697`
Xianda Sun, A. T., Philipp Gabler, & Ge, H. (2024). Juliabugs: A graph-based probabilistic programming language using bugs syntax. In *The languages for inference (lafi) workshop at the 51st acm sigplan symposium on principles of programming languages (popl 2024).*