Interactive Bayesian Probability for Learning in Diverse Populations

Zainab Attahiru
Computer Laboratory
University of Cambridge
zia21@cantab.ac.uk

Rowan Hall Maudslay
Computer Laboratory
University of Cambridge
rh635@cam.ac.uk

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

Abstract
This paper builds on work presented at PPIG 2021, "Visualising Bayesian Probability in the Kalahari" (Blackwell et al., 2021). Here, we describe a proposed interface for visualising Bayesian probability in a non-Western educational setting—in this case, a school in Nigeria. We evaluate the approach in a school workshop combining didactic and discovery methods of learning. The results offer insights for future curricula in high-school level probability, as well as an agenda for Computer Science education research, exploring how a programming perspective (through model building) might improve reasoning and intuition about probability.

1. Introduction
A paper at PPIG 2019 set out an agenda for the usability of probabilistic programming languages, including a "furthest-first" agenda that would develop new languages for use in schools of Sub-Saharan Africa (Blackwell et al., 2019). At PPIG 2021, a progress report on that programme of work described an investigation toward visualising Bayesian probability in the Kalahari (Blackwell et al., 2021)

In this paper, we extend that programme of work with a report on visualising Bayesian models in Nigeria, with education and learning as primary considerations in the design process. This builds from related work that have attempted to simplify and improve probabilistic reasoning using visualisations by employing storytelling structures (Erwig & Walkingshaw, 2013), including diagrams in live programming environments (Gorinova, 2015), and incorporating causality in the language of probability and statistics (Pearl, 1995). Pearl's work in particular, provides the foundation of the visualisation presented in this paper.

This work is motivated by previous work in computer science education research that has explored the efficacy of diagrams and storyboards in learning (Waite et al., 2016) as well as the success of visual languages such as Scratch (Resnick et al., 2009) and Sonic Pi (Aaron & Blackwell, 2013) as first programming languages. Our work investigates the potential of diagrams and graphical objects to produce equally successful outcomes in other areas such as probability education in non-Western contexts.

2. Background
2.1. Visualisations in Child Learning
Exploring external representations as a learning methodology for children goes back to education pioneers such as Froebel in the 19th century (Manches et al., 2010). Piaget for example, highlights the importance of the manipulation of concrete objects in learning via his constructivist theory (Von Glasersfeld, 1982). Papert (1980), inspired by Piaget, developed a theory of constructionism where he argues that these manipulable objects provide an effective learning tool as they are embedded in a child's cultural environment. Papert's insights have been influential in later works including those concerned with the development of digital interfaces and development environments for kids (Resnick et al., 2009; Stead, 2016). Scratch (Resnick et al., 2009) for instance, uses graphical objects in the form of legos as constructs to allow younger audiences explore programming and art creativity. This has been a successful endeavour boasting a record of 42 million users in 2021. Manches et al. (2010) have posited that the success of external representations is attributable to their ability to offload cognition by mapping processes whilst providing conceptual metaphors that allow children to relate concepts to real world experiences. This helps with knowledge transfer, an important design goal for building educational tools

(Stead, 2016).

Despite the success of Scratch, the use of legos as the foundation of the graphical representation signifies Western preferences. Thus, the transferability of the success of such visual representations to kids that are not exposed to Western culture is unknown. This provides an avenue for exploring the effect of external representations on children from varying backgrounds and the possibility of a "general" visualisation that could be discovered from a non-Western culture.

2.2. Explaining Probabilistic Models

The ubiquity of intelligent tools raises questions relating to user interaction. User in this context is not limited to the scope of the generic end-user but includes other stakeholders such as programmers and system designers. A prominent design approach to this problem has been to improve the 'explainability' of these tools by helping users build accurate mental models of system behaviour. As the underlying models of these intelligent tools are probabilistic in practice, some of the work done has sought to explain the inference process in probabilistic programs.

Erwig & Walkingshaw (2013) for example, designed a visual metaphor based on the causal structure of stories to breakdown the stages of a probabilistic modelling process. This representation serves as an explanation that maps the discrete values of a distribution to the probability of their outcome. The authors use the Monty Hall problem to buttress the intuitive nature of their representation and further provide insights into how their work could improve certain programmer-based tasks such as debugging. A related work has defined and discussed debugging principles that could guide the design of explanations (Kulesza et al., 2015). Gorinova (2015) on the other hand focused on how visual explanations could aid the probabilistic programming process. The author developed a live multiple representation development environment (MRE) that shows the Bayesian network of variables based on an infer.NET probabilistic model. Participants using the MRE were found to provide high-level descriptions of underlying probabilistic dependencies compared to those using a conventional development environment.

This has led to a promising view of graphical visualisations as an effective means of improving probabilistic reasoning. Blackwell et al. (2021) have especially discussed the potential of interactive visualisations for teaching probability in schools whilst looking at a furthest-first design outlook that is inclusive of underserved communities.

2.3. Causal Inference

Causal thinking is a cognitive principle used by humans and its deductive nature plays a significant role in scientific thought and discovery. The domain of statistics and probability, as exemplified by its popular mantra "Correlation does not imply Causation" discounts this principle, resulting in non-intuitive language for communicating concepts and paradoxes. The science of Causal Inference propounded by Pearl (1995) seeks to provide tools that mathematise the concept of causality by building on current statistical methods. This approach is gaining support in the AI community as it provides a pathway to solving learning problems using less data as well as tackling ethical problems in algorithmic design such as fairness (Kusner et al., 2017).

Pearl asserts that a purely statistical approach to AI is inadequate as it neglects the causal assumptions and beliefs that are necessary for producing intelligent behaviour. This leads him to formulate a hierarchical division of intelligence referred to as the Ladder of Causation. Each rung of the ladder has additional expressiveness compared to the rung below it, which determines the type of questions it can answer (Bareinboim et al., 2022). The rungs and their assessment of likelihood of outcomes are presented below;

- Association: The likelihood of an outcome is assessed by observed evidence. Most of traditional statistics and data-only machine learning solutions lie in this rung. An example of a question on this rung is "What is the likelihood that a toddler will be vaccinated?".

- Intervention: The likelihood of an outcome is assessed by introducing an intervention and measuring its effects. This is not the same as an assessment by conditioning on an observed value of a variable but rather by a forced manipulation of the variable. This is particularly useful in healthcare research where randomised trials may not be feasible or are ethically sensitive (Zhang et al., 2021). An example of a question on this rung is "How many people will die if the vaccine dose is doubled?".

- Counterfactual: The likelihood of an outcome is assessed based on a hypothetical scenario. The rung alludes to imagining and Pearl argues that imagining is responsible for human intelligence and thus essential for achieving artificial general intelligence. AI research has employed counterfactuals to proffer solutions in computational advertising (Bottou et al., 2013) and algorithmic fairness (Kusner et al., 2017). An example of a question on this rung is "What will happen if a person who took a vaccine decides not to take it?".

Causal Inference as a technical discipline specifically deals with answering interventional and counterfactual questions. To effectively do this, one has to be able to represent what they know (assumptions) and what they want to know (queries). Causal inference provides tools for representing these knowledge. Assumptions are expressed using two methods; mathematically using Structural Causal Models (SCMs) and diagrammatically using Causal Diagrams. These representations are complementary. Queries are expressed using Do-Calculus. This paper will be focused on providing a brief overview of SCMs and Causal Diagrams. For an in-depth coverage of Do-Calculus, refer to Glymour et al. (2016).

2.3.1. Structural Causal Models

When defining SCMs, one has to be cognisant of the two variable types in causal modelling; exogenous and endogenous variables. Exogenous variables are background variables whose causes are determined by factors outside of the model. Endogenous variables are those variables whose causes or values are known and defined within the model. Exogenous variables cannot be descendants of other variables as their causes are unknown while endogenous variables can be descendants of exogenous or endogenous or both types of variables. The relationship between child and parent variables are mapped by functions. This in turn induces strong assumptions about the nature of the relationship but provides a convenient syntax for how the distribution of these variables will change in response to external interventions (Pearl, 1995).

Definition 2.1 *A causal model can be defined as a triple of sets $\langle \mathcal{U}, \mathcal{V}, \mathcal{F} \rangle$, where $\mathcal{V}$ is a set of endogenous (observed) variables, $\mathcal{U}$ is a set of exogenous (latent) variables, and $\mathcal{F}$ is a set of functions (each one corresponding to a variable $V_i \in \mathcal{V}$), such that*

$$V_i = f(pa_i, U_{pa_i})$$

where $f \in \mathcal{F}$, $pa_i \subseteq \mathcal{V}$ are the parents of V_i, and $U_{pa_i} \subseteq \mathcal{U}$ are some independent noise variables and can be identified as the "causes" of V_i.

2.3.2. Causal Diagrams

These are directed acyclic graph visualisations of SCMs. Solid nodes in the graph structure represent endogenous variables while hollow nodes represent exogenous variables. The functional relationship between variables are represented using directed edges that connect variables. Causal diagrams aid in answering causal queries and computing causal effects in the absence of empirical data (Forney & Mueller, 2021). In addition, they provide an appropriate visual for structures that could result in statistical paradoxes such as confounding, mediation and collision.

Causal diagrams bear a close resemblance to Bayesian networks and are indeed built from the same foundation. By employing a causal structure in their construction, we propose that these diagrams could provide an alternative visualisation for representing and explaining Bayesian probabilistic models that is closer to human causal thinking. The abstract form of the graph provides a more universal visual approach as opposed to established norms of dices and coins in probability explanations that could have cultural connotations. In addition, the basic elements of causal diagrams - nodes and directed arrows

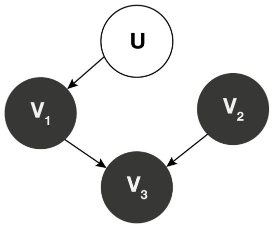
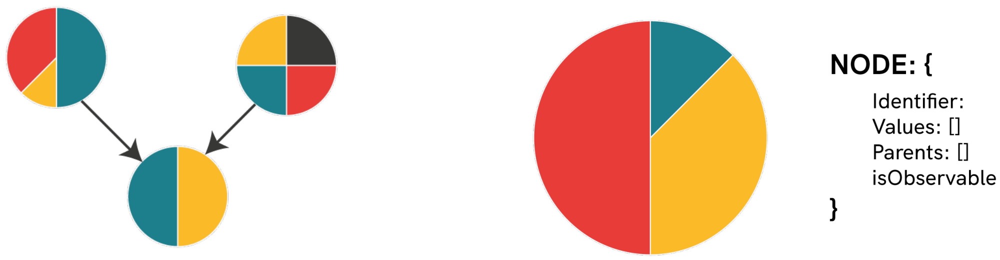


Figure 1 – A simple model represented as a causal diagram



(a) Multiple NODE *and* CONNECTOR *objects* *(b) A single* NODE *object and its attributes*

Figure 2 – A visual notation for casuality

- provide basic building blocks that are similar to those available in graphics software such Adobe Illustrator. These can be converted to manipulable graphical objects that could transform the model building process into an interactive experience.

3. Work

3.1. A Modified Visual of Causal Diagrams

Building on our decision to adopt causal diagrams to visualise Bayesian probability, we modify the graph to accommodate this objective. In our modified visual formalisation (as shown in Figure 5), we treat both endogenous and exogenous variables as a single graphical object, NODE. This represents any random variable of a typed discrete finite probability distribution. The visualisation is in the form of a pie chart, with sectors of the chart representing value-likelihood pairs. Functions are represented using a CONNECTOR object, a directed arrow between nodes similar to Figure 1 (see Figure 2a).

A NODE is defined using four attributes (as shown in JSON in Figure 2b right) : an identifier, values, parents, and a flag which marks whether it is observable. Below is a breakdown of what each of these attributes represents:

- Identifier: The identifier holds a string name assigned to a random variable. In basic formulations, the identifier could be prove to be risky for users as it places a variable naming burden. This could be simplified by including predefined options in the naming interface or alternatively, by inferring the variable name from the content of the values attribute as done in spreadsheets (Sarkar et al., 2022).

- Values: The values attribute holds the different values that a random variable can take: it is an array of tuples, each containing a name attribute (of the value) paired with a likelihood attribute.[1] In the visualisation (Figure 2b), each value corresponds to a sector of the pie-chart. The likelihood attribute specifies how likely the user assesses that an outcome is to occur and therefore holds their prior belief about the random variable. This is represented using floating point numbers (rounded to two decimal places).

  In defining this attribute, a burden is placed on users by prompting them to identify possible out-

[1]Since we are dealing with discrete probability distributions, these are countable.

comes of a random variable and their likelihood. This induces a form of premature commitment and could be classified as a hard cognitive task in instances where users don't have a solid understanding of what a random variable constitutes.[2]

- Parents: This attribute holds the parent variables of a random variable: it is an array of tuples, each containing a parent's identifier and an array specifying the effect of the parent on the variable's distribution. This defines the relationship between child and parent variables, and is represented visually using the `CONNECTOR` graphical object (a directed arrow, see Figure 2a).

- isObservable: This attribute specifies the latency of a random variable and is used to determine which variables are exogenous and which are endogenous. This allows us to represent both variable types using the `NODE` object. Setting this attribute to false places a restriction on adding parent nodes, since exogenous variables' causes are unknown (owing to the nature of the variable or the modeller's choice).

3.2. An Interface for Defining the Diagrams

To produce the visual formalisation presented in the previous section, we sketch out the design of interactive interface components that can be used to create expressions of causal dependencies. This interface induces the following interaction flow: declare variable, specify value distribution, and define relationships between variables.

3.2.1. Declaring Variables

Figure 3a shows a variable declaration component used to create "pseudo" `NODE` objects (this only specifies the identifier and observability; values and parents are left to later components). The observability of the variable is posed as a question "Can you observe it". An earlier iteration considered the question "Can you see it" as a more natural form, but this proved to be an oversimplified and narrow depiction of observation. To accommodate for the more complex idea of observation, a tooltip is included to express the different ways of observing.

In our discussion on the identifier and values attributes (§3.1), we mentioned the potential cognitive strain imposed on the user during naming. While we referred to selecting from a predefined set of variables (based on a particular domain of application such as health as per the presented design) and inferring from the values attributes as mitigations, another alternative could be the use of the "Variable Description" field to serve as prompt for the variable's identifier (denoted as "Variable" in the presented design). This serves as a guide and teaching aid whilst allowing users to retain a level of autonomy in determining what variable they want to represent.

3.2.2. Specifying Value Distributions

Figure 3b presents the distribution builder component used specify the values of a variable while ensuring that the sum of their likelihoods is equal to one. Distributions are defined from selecting between binary and range options and value names can be edited based on the user's preferences. The value names in Figure 3b serve as placeholder text and a simple explanation of how value names can be defined. The distribution of the values is represented using bar charts with draggable bars. The length of each bar corresponds to a value's likelihood. Likelihood is assigned by dragging the bar which provides both a spatial visualisation and relative depiction (to other values). Additionally, likelihoods are presented in percentages to provide an alternative representation for understanding the concept of likelihood.

Each bar is responsive to changing values in the other bars (dragging a bar results in a change of value in other bars) as a means of enforcing the sum rule of discrete probability distributions. This provides a convenient modal of interaction for binary outcomes but could be cumbersome for distributions with three or more values. This can be bypassed by disabling the responsive dragging and performing validation checks before saving. Additionally, values can be added and removed. A constraint of the number of values allowed per variable can be reinforced to allow for clarity in visualisation.

[2]This could once again be simplified using predefined options.

(a) Declaring variables *(b) Specifying distributions* *(c) Defining relationships between variables*

Figure 3 – Prototypical UI components

3.2.3. Defining Relationships

Defining relationships between parent and child variables in causal diagrams is a difficult task and requires assumptions based on knowledge of probability distributions such as normal, poisson or binomial distributions. As our objective is to employ a graphical approach suitable for younger users, we define these relationships not as functions between distributions but rather by conditioning.

In the relationship builder component (Figure 3c), we combine two bar chart representations with different interaction constraints. Each horizontal bar (to the left) represents a parent variable's values with their width corresponding to their likelihood. These are bars are selectable and selecting a bar re-renders the child variable's values (denoted by the vertical bar chart). These values are vertical draggable bars that can be used to specify the effect of a parent variable's value on the child variable's values. The text with the following template - "How does [parent variable's value] affect [child variable]" reinforces the causal notion that the parent is a determinant factor in the values of the child.

The component's construction on only defining the effect of a single parent results in the marginalisation of other parent variables. When specifying prior distributions, this is adequate. However, when posterior distributions are specified (in the form of edits), this could lead to an inference problem.

4. Workshop

To determine the efficacy of our modified visual of causal diagrams as a tool for explaining Bayesian probabilistic models, we performed a short user study with secondary school students in Nigeria.

4.1. Study Design

A web application consisting of four separate modules, representing four tasks was used as the study environment. Rather than tackle the question of the explainability of our modified causal diagrams directly, we opted to answer the more concrete question of measuring the influence of the visualisation on the participant's decision making. This goal was broken down into the following research questions:

1. Are secondary school students in Nigeria using Bayesian reasoning in their decisions?
2. Can Bayesian reasoning be introduced through an interaction with the modified causal diagram?

With these questions, we designed the study as follows - the first and fourth tasks were decision-making tests while the second and third tasks served as a short lesson and tool interaction session respectively. By comparing the results from the decision-making tests, we can investigate the effect of the visualisation on the answers provided by the participants.

To determine what type of questions will feature in the tests, we considered domains of application that will be culturally relevant to the participant group. For the questions in first decision-making test as well as the content of short lesson, we chose the COVID-19 pandemic and further narrowed the scope to a "COVID in the Market" scenario. The global impact of the pandemic and its depictions within public discourse provides a familiar space where participants can reason about events and their causal structure. For the second decision-making test, we opted for a more localised domain - tree climbing.

The inclusion of a short lesson as opposed to direct exposure (and exploration) of the visualisation is intended to create a structure that is familiar to the participant group, who are conversant with learning environments that have a didactic component in contrast to their Western counterparts. Sfard (1998) has also asserted that a mixture of both didactic and discovery methods are essential in learning.

4.2. Participants

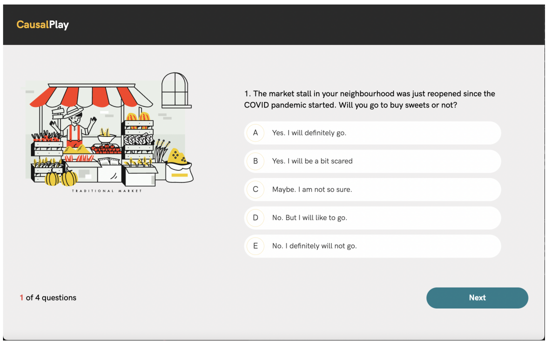
Eight participants were recruited from Government Secondary School Wuye in Abuja, Nigeria. Their prior exposure to probability is based on rote method computation of mutually exclusive events as opposed to extensive understanding of concepts such as randomness and conditioning.

4.3. Procedure

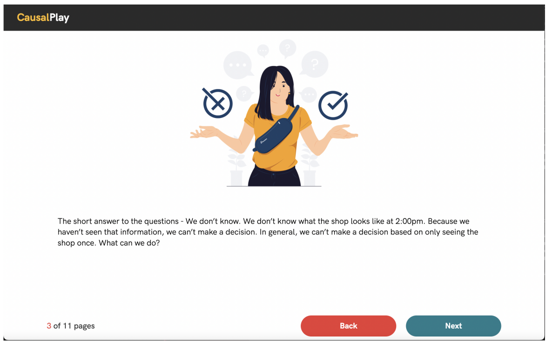
We used a within-subjects experiment design for the study. The study lasted for approximately an hour.Each participant was provided with a desktop computer to access the study environment online. The participants were required to complete the four tasks in order. During the first and fourth tasks, participants were not allowed to interact with each other in order to preserve the validity of the answers they provide. The second and third tasks were in the form of an open engagement session. Below is a breakdown of the details of each task;

Participants completed four tasks, detailed below. In the first and last tasks, students make decisions in hypothetical scenarios. In the middle tasks, they are introduced to Bayesian reasoning. The aim is to observe the influence of introducing the notions of randomisation and conditioning on the certainty of the decisions made by the participants.

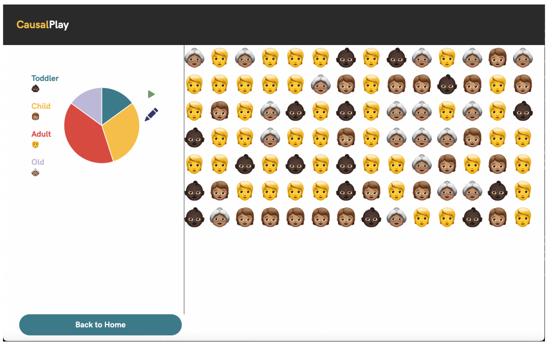
1. Task A: Participants were required to make decisions based on three hypothetical scenarios placed within the "COVID in the Market" domain. The questions include:
   - *(a) The market stall in your neighbourhood was just reopened since the COVID pandemic started. Will you go to buy sweets or not?*
   - *(b) Your friend has eyes that can see COVID. She went to the stall at 10 in the morning and said half the people she saw had COVID. Will you still go at 2 in the afternoon?*
   - *(c) It has been three (3) days since your friend went to the stall and she seems fine. Will you go to the stall today?*
2. Task B: Participants read through ten short explanations (including both textual and visual imagery) on Bayesian reasoning and causal diagrams.
3. Task C: Participants interact with a tool based on our modified causal diagram visualisation with emphasis on the `NODE` object. This was achieved by presenting a predefined variable representation and allowing students to run sampling simulations. Interactive discussions during the session prompted students to reason about changing distributions of random variables when placed in different contexts (for example when a background variable changes).The complete version of the Figure 3 was not developed for this study.
4. Task D: Participants were required to make decisions based on three hypothetical scenarios in a tree climbing scene. The questions include:
   - *(a) You belong to a group of people who love to climb one particular tree. You have been doing this for the past two years. How likely are you to break your leg from climbing a tree?*



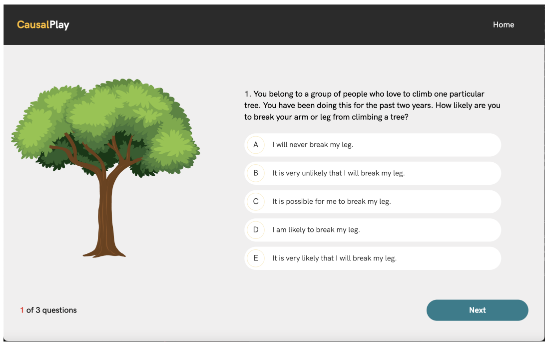
(a) Task A



(b) Task B



(c) Task C



(d) Task D

Figure 4 – Study Environment Screenshots

(b) *Your best friend recently broke both legs climbing this tree. Will you join the others to climb the tree tomorrow?*

(c) *Fatima just moved into your neighbourhood. She can run very fast. How good will she be at climbing trees?*

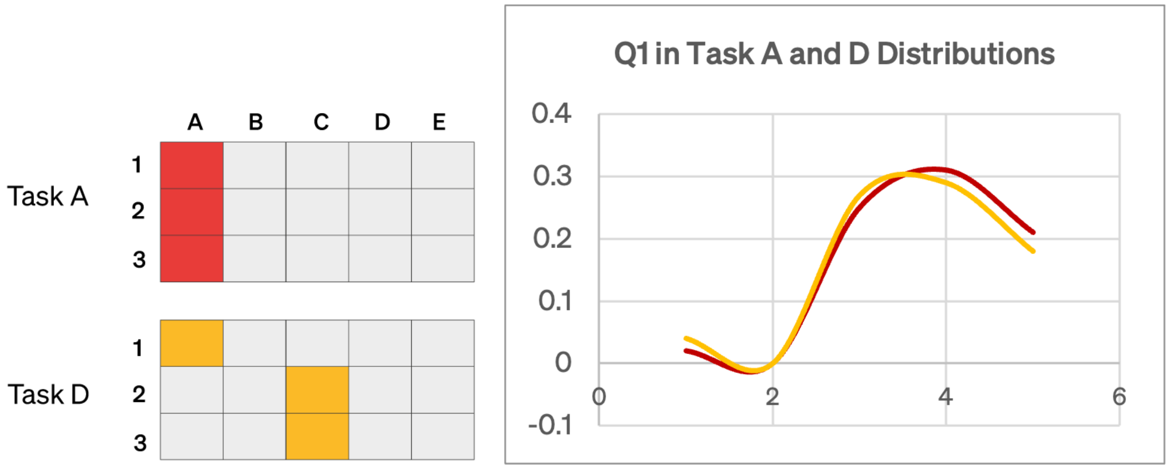
4.4. Findings and Discussion

The responses to the two decision-making tasks were collected and analysed. For each question, participants were required to give a response from five options. See Figure 5a for the map of a participant's decisions during the tasks. The table below provides a scheme for each option;

Option	Content	Scale	Description
A	Yes, I will definitely go to the market	5	Positive
B	Yes, I will be scared	4	Semi-positive
C	I am not sure	3	Uncertain
D	No, but I want to go	2	Semi-skeptic
E	No, I will absolutely not go	1	Skeptic

The standard deviation of both sets of responses was calculated to determine their distribution in order to investigate the questions posed in the study's design. The expectation is that the employment of Bayesian reasoning will induce a normal distribution of responses with most responses in the uncertain region (corresponding to 3 on the scale). This is because the hypothetical scenarios provide little evidence to cause a decision shift towards the extreme ends of the range. We provide discussions to the study research questions below:

Are secondary school students in Nigeria using Bayesian reasoning in their decisions? The data shows that this is not the case. Responses to the first decision-making test were skewed to extreme ends of the



(a) A Participant's Response Map *(b) The distributions of responses to questions*

Figure 5 – Analysing study findings

range across all questions. In the second question for instance, the uncertain ("I and skeptic choice were not selected by any of the participants. However, this result is to not surprising. There is a wide range of research that has been undertaken that shows that people find it difficult to think in terms of probabilistic quantities.

Can Bayesian reasoning be introduced through an interaction with the modified causal diagram? The data once again shows that this is not the case. Similar to the first decision-making test, the second test's responses are also skewed towards extreme ends. The distribution of the responses to the first questions in both tests exhibit a striking resemblance as depicted by Figure 5b. This result could have arisen due to size of the participant group, which represents a small (and perhaps biased) sample of the target population. Furthermore, research in learning are in the format of studies that measure effects over a longer period of time. The expectation that a short study could produce a similar effect is unrealistic.

These results raise new questions not the least of which are those surrounding the validity of the decision-making tests. One worth noting is how the effect of different biases on different decision scenarios could be taken into account when designing future user studies.

5. Conclusion and Future Work

n this paper, we argue that representations and demonstrations of Bayesian probability can be redesigned to tackle problems related to inclusivity in education. We highlight previous work undertaken on the use of visualisations in child learning and in representing probability models. This lead us to modify Pearl's causal diagrams to include additional handles for specifying Bayesian probability models with interactivity as a key design consideration.

A study was conducted in Nigeria using a simplified form of the modification and the explainability of the visualisation was assessed via its effect on the participant's decisions. Findings from this study were limited by a variety of factors including the scale of the software implementation, study length, and participant sample size. Future work on this project will focus on expanding the functionality of the software environment to allow for more dynamic model building, taking inspiration from graphics editing software such as Adobe Illustrator. This will be accompanied by longitudinal user studies to determine the efficacy of the visualisation and interaction environment as learning tools, as well as the selection of adequate metrics for measures of interest such as retention and knowledge transfer.

References

Aaron, S., & Blackwell, A. F. (2013). From sonic pi to overtone: Creative musical experiences with domain-specific and functional languages. In *Proceedings of the first acm sigplan workshop on func-*

tional art, music, modeling amp; design (pp. 35–46). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2505341.2505346 doi: 10.1145/2505341.2505346

Bareinboim, E., Correa, J. D., Ibeling, D., & Icard, T. F. (2022). On pearl's hierarchy and the foundations of causal inference. *Probabilistic and Causal Inference*.

Blackwell, A. F., Bidwell, N. J., Arnold, H. L., Nqeisji, C., Kunta, K., & Ujakpa, M. M. (2021). Visualising bayesian probability in the kalahari. *Psychology of Programming Interest Group (PPIG)*.

Blackwell, A. F., Church, L., Erwig, M., Geddes, J., Gordon, A., Gorinova, M., ... others (2019). Usability of probabilistic programming languages. *Psychology of Programming Interest Group (PPIG)*.

Bottou, L., Peters, J., Quiñonero-Candela, J., Charles, D. X., Chickering, D. M., Portugaly, E., ... Snelson, E. (2013). Counterfactual reasoning and learning systems: The example of computational advertising. *Journal of Machine Learning Research*, *14*(65), 3207–3260. Retrieved from http://jmlr.org/papers/v14/bottou13a.html

Erwig, M., & Walkingshaw, E. (2013). A visual language for explaining probabilistic reasoning. *Journal of Visual Languages Computing*, *24*(2), 88-109. Retrieved from https://www.sciencedirect.com/science/article/pii/S1045926X13000025 doi: https://doi.org/10.1016/j.jvlc.2013.01.001

Forney, A., & Mueller, S. (2021). Causal inference in ai education: A primer.

Glymour, M., Pearl, J., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

Gorinova, M. I. (2015). Interactive development environment for probabilistic programming.

Kulesza, T., Burnett, M., Wong, W.-K., & Stumpf, S. (2015). Principles of explanatory debugging to personalize interactive machine learning. In *Proceedings of the 20th international conference on intelligent user interfaces* (p. 126–137). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2678025.2701399 doi: 10.1145/2678025.2701399

Kusner, M. J., Loftus, J., Russell, C., & Silva, R. (2017). Counterfactual fairness. In I. Guyon et al. (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper/2017/file/a486cd07e4ac3d270571622f4f316ec5-Paper.pdf

Manches, A., O'Malley, C., & Benford, S. (2010). The role of physical representations in solving number problems: A comparison of young children's use of physical and virtual materials. *Computers Education*, *54*(3), 622-640. Retrieved from https://www.sciencedirect.com/science/article/pii/S0360131509002632 (Learning in Digital Worlds: Selected Contributions from the CAL 09 Conference) doi: https://doi.org/10.1016/j.compedu.2009.09.023

Papert, S. A. (1980). *Mindstorms: Children, computers, and powerful ideas*.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, *82*(4), 669–688. Retrieved 2022-06-03, from http://www.jstor.org/stable/2337329

Resnick, M., Maloney, J., Hernández, A., Rusk, N., Eastmond, E., Brennan, K., ... Kafai, Y. (2009, 11). Scratch: Programming for everyone. *Communications of the ACM*, *52*, 60-67.

Sarkar, A., Ragavan, S. S., Williams, J., & Gordon, A. D. (2022). End-user encounters with lambda abstraction in spreadsheets: Apollox2019;s bow or achillesx2019; heel? In *2022 ieee symposium on visual languages and human-centric computing (vl/hcc)* (p. 1-11). doi: 10.1109/VL/HCC53370.2022.9833131

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational researcher*, *27*(2), 4–13.

Stead, A. G. (2016). Using multiple representations to develop notational expertise in programming. *University of Cambridge, Computer Laboratory, Technical Report*(UCAM-CL-TR-890).

Von Glasersfeld, E. (1982). An interpretation of piaget's constructivism. *Revue internationale de philosophie*, 612–635.

Waite, J., Curzon, P., Marsh, W., & Sentance, S. (2016). Abstraction and common classroom activities. In *Proceedings of the 11th workshop in primary and secondary computing education* (pp. 112–113). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/2978249.2978272` doi: 10.1145/2978249.2978272

Zhang, W., Ramezani, R., & Naeim, A. (2021). *Causal inference in medicine and in health policy, a summary.* arXiv. Retrieved from `https://arxiv.org/abs/2105.04655` doi: 10.48550/ARXIV.2105.04655