Bringing math into politics: the Decision Theatre Lab and the influence of model knowledge on the interpretation of stochastic simulation results

Christoph Lieben[1] and Sinah Gürtler[2]

[1]Humboldt University of Berlin, Germany; c.lieben@hu-berlin.de

[2]Free University of Berlin, Germany

Results of stochastic models in public media are often misunderstood, not least because the uncertainties of stochastic modelling are not sufficiently considered in classical mathematics teaching. This paper first presents a learning environment that confronts students with different forms of mathematical uncertainty: the Decision Theatre Lab. In it, students mathematically explore a professional epidemic model and use its stochastic results for political decision-making. We then present an experimental study that we conducted using this format to investigate how specific model knowledge influences the interpretation of simulation results. The results show that students with specific model knowledge are significantly better at accurately interpreting the stochastic variations in the simulation results than students who had to interpret the results without prior knowledge of the underlying model. The latter, on the other hand, rely more often on contextual knowledge to interpret the model results.

Keywords: Model knowledge, mathematical literacy, stochastic modelling, model uncertainty

Introduction: Mathematics education in the era of vague news

Public media regularly report on the results of scientific modelling, in particular stochastic modelling of climate change. Statements such as “Climate change made US and Mexico heatwave 35 times more likely” (Brosnan, 2024) are prone to misinterpretation, driven by a lack of statistical knowledge and misconceptions about mathematical models. The fact that the underlying models are often inaccessible to readers does help at all.

Gal and Geiger (2022) have provided a detailed taxonomy of the *Statistical and Mathematical Products* (StaMPs) appearing in public media in the context of the COVID-19 pandemic. They conclude that the interpretation of these STaMPs requires not only an academic understanding of probability but also a literate handling of more subtle and unquantifiable forms of uncertainty that arise, for example, in the case of inadequate data or when making modelling assumptions. They diagnose the need for learning environments that consider these forms of uncertainty.

Against this background, we will first introduce the *Decision Theatre Lab*: a science communication format that enables students to use a scientific epidemic model to make political decisions in a fictitious but reality-based setting. We describe the structure of a Decision Theatre Lab, analyse which forms of uncertainty are inherent in the occurring StaMPs and which aspects of mathematical literacy are important for their interpretation. We then present a case study that used the Decision Theatre Lab framework to investigate how specific knowledge of the model influences students’ interpretation of its stochastic results.

Exposing students to uncertainty – a Decision Theatre Lab on infection control

A Decision Theatre Lab – literally and practically – consists of two components: a *Decision Theatre* and a *School Lab*. A Decision Theatre is an IT-supported science communication format in which participants can simulate joint decision-making on various socially relevant topics, guided by the interaction with a scientific model and its simulation results. Decision Theatres have been developed on a variety of topics, from forest planning (Boukherroub et al., 2018) to mobility transition (Wolf et al., 2023) – or infection control (Wiebe et al., 2024), which students attended at Zuse Institute Berlin as seen in Figure 1 and which provides the framework for the study presented here.

In this Decision Theatre, participants take on the role of politicians who have to decide on non-pharmaceutical infection control measures at the beginning of a fictitious epidemic. They can draw on the results of GERDA, a GEoReferenced Demographic Agent-based model for the spreading of COVID-19 (Goldenbogen et al., 2022). GERDA simulates the location, the infection probability and the health status of about 12,000 residents of a small town and provides SIRD trajectories for different scenarios. Due to the stochastic modelling, the simulations can differ significantly from each other, even with the same initial conditions. Therefore, the model simulates each scenario almost 100 times and displays the results as a set of curves for the number of susceptible, infected, recovered and deceased. The results should be interpreted as probability distributions that show different degrees of statistical dispersion depending on the scenario. The simulation results depicted in Figure 2, for example, show that dispersion increases the later a potentially imposed lockdown comes into force. This indicates that the prediction accuracy of the model decreases with a later lockdown. Phenomena like these typically emerge in stochastic models (Goldenbogen et al., 2022).

Figure 1: Students in a Decision Theatre

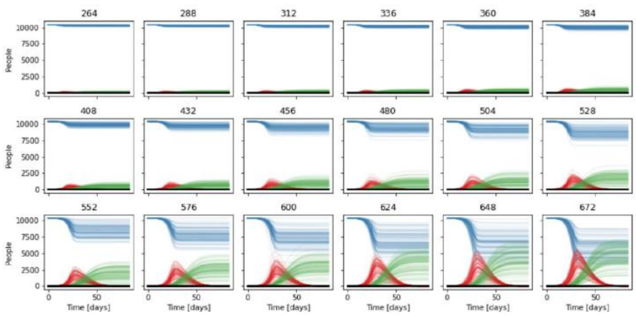

Figure 2: GERDA results for different start times (above graphs in h) of a lockdown

To support their political decision-making, students can simulate three scenarios with different political measures. In the further course of the Decision Theatre, they also reflect on the role of the model for the discussion and evaluate the informative value of the simulation results.

It is important to note that although the participants in a Decision Theatre interact with a mathematical model, this model itself remains a black box for the most part. To put the format in the context of mathematics education, we have therefore developed an optional School Lab that teaches students the mathematical background and the internal structure of GERDA, with a particular focus on the implemented probabilistic concepts: The students derive the probability function for the transmission of the virus and use it to calculate transmission probabilities in various scenarios. They also use school knowledge of conditional probabilities to determine the probabilities of change in the health status of different people. These calculations, however, are all done once and with pen and paper. It is only after the School Lab in the Decision Theatre that students are confronted with the phenomenon of multimodal distributions resulting from repeated simulations.

The role of model knowledge in the interpretation of stochastic simulation results

The students' use of GERDA exemplify that the interpretation of Statistical and Mathematical Products (StaMPs) produced by stochastic models is characterised by two types of uncertainty:

- First, uncertainty arises from stochastic modelling itself, particularly from the probability functions and distributions; we refer to this simply as probability. Probability is quantifiable, and its occurrence in the model is intended. It is a basic concept of statistical education, and dealing with it is included in any description of statistical literacy, for example in the influential one by Gal (2002). The correct description and interpretation of GERDAs results, for example, requires basic ideas about the dispersion of probability distributions.
- Secondly, uncertainty is inherent in any form of mathematical modelling. It arises from the fact that models are abstractions, characterised by simplifications and based on certain assumptions. Model uncertainty is a collective term for the reasons that lead to unintended differences between the model and reality and is often difficult to quantify (Thompson & Smith, 2019). An awareness of model uncertainty is therefore an implicit prerequisite for a comprehensive modelling competence, which also includes the evaluation of the validity and scope of given modes, as described by (Niss & Højgaard, 2019), for example. Research conducted during earlier iterations of the Decision Theatre Lab has shown that students tend to explain deviations between model forecasts and actual developments with contextual knowledge rather than with model uncertainty, indicating a low level of awareness of model uncertainty (Lieben, 2023).

Now, one can assume that students depend on appropriate statistical content knowledge and sufficiently developed model competence to accurately interpret such StaMPs – in addition to some general mathematical competencies such as understanding graphs and a positive mathematical identity, as Heyd-Metzuyanim et al. (2021) note. However, the observation of some unsuccessful StaMP interpretations in Decision Theatres without the prior School Lab led us to assume that this framework, as shown in Figure 3, can be extended to include another construct, namely, the specific model knowledge that is acquired in the School Lab.

But since a subject's prerequisites for interpreting (COVID-19-related) StaMPs in the media are usually formulated in terms of statistical literacy and modelling skills – as in Gal & Geiger (2022) or Heyd-Metzuyanim et al. (2021), the influence of specific model knowledge on the interpretation of the model results has so far been underexposed. We, therefore, used the Decision Theatre Lab on infection control to conduct a case study approaching the question: *How does specific knowledge of a model influence the interpretation of its results?*

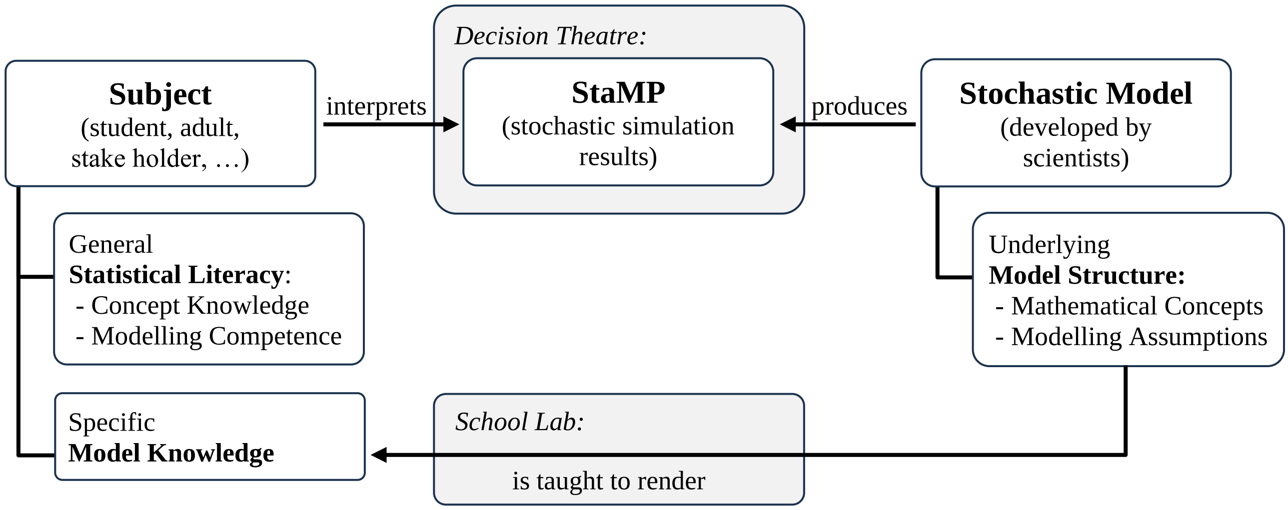

Figure 3: Conceptual framework of a StaMP interpretation in the context of a Decision Theatre Lab

Studying the influence of model knowledge – research design and implementation

To determine the influence of the specific model knowledge acquired in the School Lab on the interpretation of the StaMPs in the Decision Theatre, we conducted an experimental study with the School Lab as an intervention. We defined two groups of students: The experimental group, who first visited the School Lab and then the Decision Theatre, and the control group, who first visited the Decision Theatre and only after that visited the School Lab.

To collect data on how the students processed the stochastic model results in the Decision Theatre, we used the participant observer method according to Weischer and Gehrau (2017) while the students were working on a collaborative group task. For this task, each group received the results shown in Figure 2 from simulation runs with different start times of a lockdown and was asked to interpret the data by completing the sentence "The later the start time of the lockdown, the...". We recorded the group's communication with microphones and analysed the audio transcripts using quantitative and qualitative methods for comparing the experimental group and control group.

Between June 2023 and April 2024, a total of $N = 89$ students from the 10th grade onwards participated in the study. The students were between 14 and 20 years old, and the average grade was 2,31. Of the students, 70 first attended the School Lab and then the Decision Theatre and were thus assigned to the experimental group. The control group, which first attended the Decision Theatre and then the School Lab, comprised 19 students. We randomly assigned working groups. A total of 24 groups were analysed, with 19 assigned to the experimental group and five to the control group.

The data was analysed using a mixed-methods approach, which enables both explorative and comprehensive as well as multi-layered perspectives on the research complex (Kelle, 2019). The material for the analysis consisted of transcripts of the audio recordings and observation protocols. We processed this data according to Mayring's content analysis (Mayring, 2019).

First, we categorised the material inductively, with a particular focus on statements that allowed conclusions to be drawn about the students' statistical concept knowledge and awareness of model uncertainty. To code the data, we used MAXQDA (VERBI Software, 2023). The obtained codes were quantitatively evaluated to carry out a correlation analysis. We calculated the mean values in each category for the experimental and the control group and tested for significance using Student's t-test. In the next step, we evaluated the transcripts qualitatively using the categories specified in Table 1.

Table 1: Coding manual for the audio-transcribed student interactions

<table>
<tr><th></th><th>Code</th><th>Definition</th><th>Example</th></tr>
<tr><td rowspan="3">Statistical Concept Knowledge</td><td>Description of dispersion</td><td>Statements that verbalise the dispersion of the graphs</td><td>"The more people get infected, the greater the dispersion."</td></tr>
<tr><td>Correct interpretation as probability distribution</td><td>Statements in which the dispersion of the curves is interpreted as a probability distribution</td><td>“The later the lockdown, the more difficult it is to predict precisely.”</td></tr>
<tr><td>Deterministic Misconception</td><td>Statements that provide incorrect non-stochastic interpretations of the results</td><td>“The later the lockdown starts, the more infections there will be.”</td></tr>
<tr><td>Awareness of Modelling Uncertainty</td><td>Contextual Knowledge</td><td>Statements based on contextual knowledge that is related to the model results</td><td>“If no one is sick anymore due to the lockdown, then no one can get infected. So, the earlier the lockdown starts, the better.”</td></tr>
</table>

Results

The quantitative analysis revealed two significant correlations for the inductively selected categories. In the domain of statistical concept knowledge, 50% of the students in the experimental group correctly described or interpreted the model results. This number was only 20% in the control group. In the domain of awareness of model uncertainty, the number of context references per minute and person was significantly lower in the intervention group (0.13) than in the control group (0.24).

We then proceeded with a qualitative analysis of the answers. We closely examined how the tasks were solved, what led to correct understanding, and what characterised the misconceptions. To illustrate this analysis, we present two exemplary cases below, which show both a typical path to correct interpretation and a widespread misconception in detail.

The following example from the control group illustrates a common statistical misconception, that is then made plausible by contextual knowledge.

Olivia:	The later the lockdown starts, the faster the infection rate increases, perhaps?
[…]	
Olivia:	Because here it has a higher increase…
[…]	
John:	Yes, but why? Because it's winter?

At the beginning of the conversation, Olivia constructs a deterministic relationship between the start time of the lockdown and the number of infected people. John then tries to explain the model behaviour with real-world contextual knowledge by asking: "But why, because it is winter?". This statement illustrates a cognitive escape from the actual mathematical task. The real-world reference not only serves the students to validate the graphs, but the students also consider it necessary to interpret the model results.

In contrast, the example from the experimental group shows a typical dynamic within a conversation in which the students move from the same type of misconception to a correct interpretation.

James:	The later the time of the lockdown, the greater the number of infected persons.
Emily:	The higher the peak of infected people.
[…]	
James:	The more infected there are.
[…]	
Emily:	But look. It's also getting thicker. So, the individual lines here no longer agree about what's happening.
[…]	
William:	Yes, that's just a probability.
[…]	
William:	So, the later, um, a lock-. The later-.
Emily:	The more difficult it is to predict what will happen.

James initially postulates a causal relationship between the lockdown's start time and the number of infected persons. Emily expands on this assumption by referring to the peak of the infection curve. However, both statements are incorrect as they imply a deterministic relationship and disregard the stochastic nature of the underlying model.

A correction process begins as Emily notices that "the lines are becoming thicker," indicating the dispersion of the curves. This observation marks the start of questioning the deterministic understanding and leads to a reconsideration of the previously assumed cause-and-effect relationship. As the discussion progresses, the participants become more aware of the probabilistic nature of the model. This awareness is reflected in the statement "that's just a probability," which explicitly acknowledges the connection between the dispersion of the curves and the induced probability distribution. The introduction of the term “probability” then enables the group to move on from the mere description of the dispersion to its correct interpretation. They discern that stochastic models do not provide fixed predictions but rather depict a range of possible scenarios with different probabilities.

Discussion: How could model knowledge affect StaMP interpretation?

The fact that the students in the control group interpreted the StaMPs presented in the Decision Theatre more precisely overall could be related to the fact that model knowledge to some extent interferes with both the domain of statistical literacy and that of mathematical modelling competence.

On the one hand, the use of stochastic concepts and terms for reproducing the model components in the School Lab seems to motivate some students in the experimental group to use the same concepts and terms to interpret the dispersed graphs in the Decision Theatre – which in this case leads them to correct conclusions. Knowledge of the model thus leads to a better understanding of the model results, since the concepts used to describe the model reappear in the interpretation of its results. On the other hand, as students in the School Lab reproduce the structure of the model, they have to follow some of the underlying modelling assumptions. This may contribute to an increased awareness of model uncertainty. The alternation between reality and the mathematical model during the reconstruction process – which is typical for stochastic modelling (Kelter et al., 2023) – could help students to distinguish more precisely between these two worlds. Thus, they are significantly less tempted to use extra-mathematical contextual knowledge from reality to explain mathematical results.

It is not clearly assessable, however, whether the model knowledge has really improved the interpretation of the simulation results in a direct way or whether the *teaching* of this knowledge in the School Lab has rather led to a promotion of statistical literacy and development of modelling competence, which in turn has led to an improved interpretation. But in any case, knowledge of the structure, assumptions and use of a professional model offers students at least a glimpse of an unfamiliar, professional "mode" of mathematical modelling (Frejd, 2014).

Finally, it should be stated that many modern computer models are obviously mathematically too complex to be reconstructed in a School Lab. The promotion of model-independent statistical literacy and modelling skills therefore remains indispensable.

Acknowledgment

This work was supported by the Berlin University Alliance (BUA) [212_ExperLab_LR_3] and the Berlin Mathematical School (BMS).

References

Boukherroub, T., D'amours, S., & Rönnqvist, M. (2018). Sustainable forest management using decision theaters: Rethinking participatory planning. *Journal of Cleaner Production, 179,* 567–580. https://doi.org/10.1016/j.jclepro.2018.01.084

Brosnan, G. (2024, June 20). Climate change made US and Mexico heatwave 35 times more likely. *BBC*. https://www.bbc.com/news/articles/czvvqdg8zxno

Frejd, P. (2014). *Modes of Mathematical Modelling: An analysis of how modelling is used and interpreted in and out of school settings.* Linköping University Electronic Press. https://doi.org/10.3384/diss.diva-103689

Gal, I. (2002). Adults' statistical literacy: meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1–25. https://doi.org/10.1111/j.1751-5823.2002.tb00336.x

Gal, I., & Geiger, V. (2022). Welcome to the era of vague news: a study of the demands of statistical and mathematical products in the COVID-19 pandemic media. *Educational Studies in Mathematics*, *111*(1), 5–28. https://doi.org/10.1007/s10649-022-10151-7

Goldenbogen, B., Adler, S. O., Bodeit, O., Wodke, J. A. H., Escalera Fanjul, X., Korman, A., Krantz, M., Bonn, L., Morán Torres, R., Haffner, J. E. L., Karnetzki, M., Maintz, I., Mallis, L., Prawitz, H., Segelitz, P. S., Seeger, M., Linding, R., & Klipp, E. (2022). Control of COVID-19 outbreaks under stochastic community dynamics, bimodality, or limited vaccination. *Advanced Science*, 9(23), 2200088. https://doi.org/10.1002/advs.202200088

Heyd-Metzuyanim, E., Sharon, A. J., & Baram-Tsabari, A. (2021). Mathematical media literacy in the COVID-19 pandemic and its relation to school mathematics education. *Educational Studies in Mathematics*, *108*(1–2), 201–225. https://doi.org/10.1007/s10649-021-10075-8

Kelle, U. (2019). Mixed Methods. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* [Handbook Methods in Empirical Social Research] (pp. 159–172). Springer VS. https://doi.org/10.1007/978-3-658-21308-4_10

Kelter, R., Schnurr, A. & Spies, S. (2023). A toolbox to demystify probabilistic and statistical paradoxes. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1212419

Lieben (2023). How working with a professional model affects students' views on the validity of simulation results. In Drijvers, P., Csapodi, C., Palmér, H., Gosztonyi, K., & Kónya, E. (Eds.). (2023). *Proceedings of the Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13)*. Alfréd Rényi Institute of Mathematics and ERME.

Mayring, P. (2019). Qualitative Inhaltsanalyse [Qualitative Content Analysis]. In U. Flick, E. v. Kardorff & I. Steinke (Eds.): *Qualitative Forschung: Ein Handbuch* [Qualitative Research: A Handbook] (pp. 468 – 475). Rowohlt.

Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. *Educational Studies in Mathematics*, *102*(1), 9–28. https://doi.org/10.1007/s10649-019-09903-9

Thompson, E. L., & Smith, L. A. (2019). Escape from model-land. *Economics*, *13*(1). https://doi.org/10.5018/economics-ejournal.ja.2019-40

VERBI Software. (2023). MAXQDA 2022 [computer software]. Berlin, Germany: VERBI Software. Available from maxqda.com.

Weischer, C. & Gehrau, V. (2017). *Die Beobachtung als Methode in der Soziologie* [Observation as a Method in Sociology]. UVK Verlag. https://doi.org/10.36198/9783838548661

Wiebe J., Lieben C. & Gretenkord E. (2024, September 3–6.). *A DecisionTheatreLab on infection disease spreading: Connecting school mathematics to societal challenges in a digital environment.* 4th Mathematics Education in the Digital Age Conference (MEDA), Bari, Italy.

Wolf, S., Fürst, S., Geiges, A., Laublichler, M., Mielke, J., Steudle, G., Winter, K & Jaeger, C. (2023). The Decision Theatre Triangle for societal challenges—An example case and research needs. *Journal of Cleaner Production*, *394*, 136299. https://doi.org/10.1016/j.jclepro.2023.136299