



Periodiek

Recurring at regular intervals Issue 2023-1



Cola Experiment

Can a scientific experiment finally put an end to the timeless debate between Coca Cola and Pepsi? Do they really taste as different as people claim they do? There is a way to find out and at its core lies probability theory.

Perio interview: Roland van der Veen

First year mathematics students have already met Roland van der Veen in the course Sets and Numbers, while older students are probably more familiar with his work on topology. Along with representation theory, geometry and physics, it constitutes the majority of his research. With this interview, we offer an insight into his life - both professional and personal.



Easy Bulgarian Banitsa

Are you interested in Eastern European cuisine? You've come to the right place! We present to you one of the easiest and most popular dishes Bulgaria has to offer. Thin pastry sheets filled with traditional white cheese, baked until crispy and golden-brown. Make sure to try it at home!

From the Board	4
The Cola Experiment	5
Perio Interview: Roland van der Veen	10
Humans vs. ChatGPT	16
Recipe	21
Brainwork	22

From the Editor in Chief

If you are reading this on a physical copy, you'll notice this issue of the *Periodiek* is slightly different than you might be used to. More compact and in an envelope. The keen reader will spot that this is because we have a different printer than we used to and these changes are to stay. If you're reading this online, thanks for being green!

In this rather Mathematical issue, we interviewed the adored lecturer Roland van der Veen and have not one, but two research articles by the honours college consortium.

Unfortunately, we don't have an exchange article this issue. But, we do get to meet the secretary of the 65th board and have a Fermi problem in order to appease the physicists and astronomers for the abundance of math in this issue.

Do you have an interesting thesis or other work you want to contribute to the *Periodiek*? Reach out to us by email!

Robert Mol

Editors

Lucía Álvarez Hereda,
Margherita Bonan,
Vedang Sumbre,
Robert Monden,
Yoana Savova,
Yagmur Yetkin,
Robert Mol,

Authors

L. Álvarez Hereda,
M. Grzegorzczuk,
R. Monden,
S. Olteanu,
Y. Savova,
R. Mol

Advertisers

ASML (p. 14)

Advertise? Contact us
at bestuur@fmf.nl

Print run 50 pieces

Press shop copy76.nl

ISSN 1875-4546

The Periodiek

is a magazine from the Fysisch-Mathematische Faculteitsvereniging and appears three times per year. Previous issues can be found at perio.fmf.nl. The board of editors can be reached at perio@fmf.nl.



From the Board

Secretary

AUTHOR: S. OLTEANU

Hello everyone! My name is Stefania Olteanu, and for this year, I will be the secretary of the FMF. We all have been board for less than 2 months, and you all might still get to know us. Because of this, I am passing on a message from the 65th board: Hello, welcome, and stick with us throughout this year!

To some extent, I can't fathom the fact that I am part of the board yet. It was a very impulsive decision, made on the day of the candidate announcement of my fellow board members. The days preceding this event, people were making guesses on the whiteboard of who applied for which position, and my name appeared at least 7 times under "Secretary". I guess the members knew better than I did, since it ended up happening less than a week later. And I can gladly tell you all that I am very happy with my choices!

However, if you think secretary was my first choice of role, you would be wrong. Surprisingly, the more time passes, the more I realise it is definitely the one that suits me best. I was then quite taken aback when I started to notice the odd similarities between each of us and our direct predecessors, not necessarily work related. They might not seem very obvious at first, but once you get to know all 11 of us, I think they become noticeable.

I believe I am now deviating from the subject, so let's rewind. The past two months have been very eventful and full of unexpected hardships, but also beautiful moments and connections. In such little time, I have learned more than I could ever hope for, both about myself and Femke, Rolf, Andrada and Roy. To me, we seemed like an unlikely group of people, and I think we were a bit surprised at how well we worked together. We've managed, even while splitting the Intern function, to successfully organise (without jumping at each other's throats) many events that resembled that sense of community the FMF is known and loved for. Taking on these intern tasks that we were not expecting, was definitely a big change in how I saw this year going for us. However, as I expected, I do enjoy managing committees (when they cooperate) and helping them whenever they are in need.

My favourite event so far was the Pottery evening, which resulted from us brainstorming relaxing activities, and me remembering what I do when I procrastinate studying for exams. It turns out, a lot of people enjoy chill, de-stressing activities from time to time! But I am generally very excited for most of the things we plan on doing from now on, from the smallest changes, to bringing old committees back to life and revitalising the association.

Contrary to my usual nature, I am often the one opening the room. If you come by before your 9 or 11 am lecture, you have a very high chance of finding me in front of my computer, playing metal music and drinking cinnamon tea, as I am at this very moment. Hope to see you around!

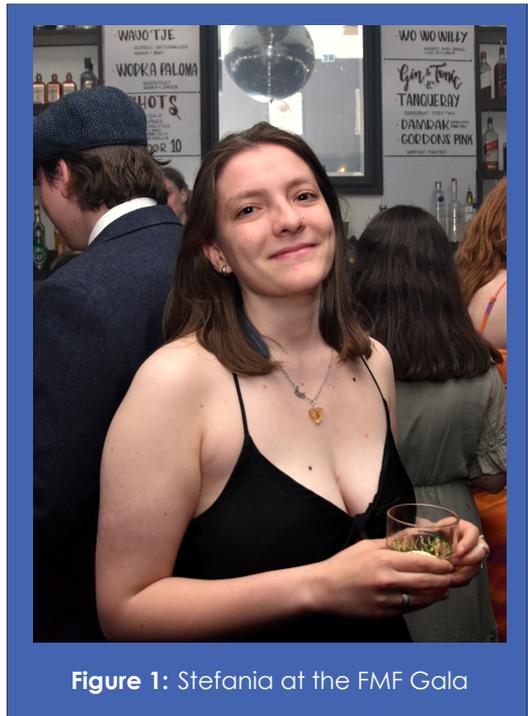


Figure 1: Stefania at the FMF Gala

The Cola Experiment

AUTHOR: M. GRZEGORCZYK AND THE HONOURS COLLEGE CONSORTIUM 2021/2022

To test whether Cola soft drink brands differ in their tastes, we organized and performed two sensory experiments, and we analyzed the experimental results statistically. In the 1st experiment, we compared the classic Cola of the two most popular brands: ‘Pepsi Cola’ and ‘Coca-Cola’. In the 2nd experiment, we compared the calorie-reduced products: ‘Light’ and ‘Zero Sugar’, both produced by Coca-Cola. In this paper, we describe the experimental setup and we report the results of our statistical analyses.

Triangle sensory tests

In a sensory experiment of the type ‘triangle test’, each participant gets three samples to taste, where two samples are from the same product and the third sample is from another product. A coin is flipped to decide which of the two products appears twice, and the arrangement of the three samples is in the form of a triangle. Thereby to avoid potential biases, the three samples are randomly assigned to the three corner points of the triangle. When the experiment has been set up, the participant enters the room and is asked to identify the deviating sample. Of course, it is ensured that the samples are presented in a neutral way, so that only sensory impressions help to distinguish them. The participant is allowed to taste each sample multiple times and in any desired order. Eventually, even if undecided, the participant has to nominate one of the three samples as a (potential) outlier. In a triangle test, when not tasting any differences or when just guessing, the probability of nominating the true (deviating) sample can be assumed to be equal to $p = \frac{1}{3}$.

Our experimental setup

In our two experiments, we served the three samples in disposable paper cups (made for coffee), and we ensured that all samples had exactly the same fridge-cooled temperature. For practical purposes, we deviated from the triangle arrangement, and we instead placed the three samples in a row. During the experiment ‘crackers’ were offered to the participants, so that they could neutralize their tastes. The participants were voluntary 2nd and 3rd year Bachelor students following the Honours College course on ‘Practical Statistical Hypothesis Testing’ (PSHT) in block 2b of the academic year 2021/2022. All 30 enrolled students were invited to participate in both experiments. $n = 27$ students participated in the 1st experiment (Pepsi Cola vs. Coca-Cola), and $n = 22$ students participated in the 2nd experiment (Light vs. Zero Sugar), both produced

by Coca-Cola). We note that 19 out of 30 students participated in both experiments. In the absence of taste differences, we would expect $e = n/3 = 9$ participants to be correct in the 1st and $e = n/3 = 7.333$ to be correct in the 2nd experiment.

Figures 2 and 3 show the products under comparison and three paper cups.



Figure 2: Coca-Cola vs. Pepsi Cola; both classic Colas were taken from 1,5l plastic bottles. Between the bottles stand three paper cups in which the samples were provided.



Figure 3: Light Cola vs. Zero Sugar Cola; both produced by Coca-Cola and taken from 0,3l cans.

Research questions

In both Cola experiments, our research questions were:

1. First, check whether there is statistical evidence that the two products can be distinguished. That is, perform a statistical test to check whether the probability p ($0 \leq p \leq 1$) for identifying the deviating product is significantly greater than $p = 1/3$.
2. Give an indication which fraction θ ($0 \leq \theta \leq 1$) of people (participants) can distinguish the two products by their tastes and provide a confidence interval for this fraction θ .
3. Screen whether there are any significant group-specific differences. For example, check for gender-specific differences and check whether regular Cola drinkers perform better.

We note that there is a difference between the probability p of nominating the right outlier and the probability that a person can taste the difference; please see below.

To address our research questions we apply the following statistical methods:

1. We perform exact Binomial tests to test the null hypothesis $H_0 : p = \frac{1}{3}$ against the alternative hypothesis $H_1 : p > \frac{1}{3}$. We perform the tests to the level $\alpha = 5\%$, and we report the p-values.
2. We employ the asymptotic efficiency of the Maximum Likelihood estimator to derive an asymptotic confidence interval for p , and we transform it into a confidence interval for the fraction θ .
3. Finally, we perform various χ^2 tests for independence to screen for any interesting significant group-specific differences.

Statistical tests and p-values

Loosely speaking, in any statistical test one checks whether there is evidence against a so-called null hypothesis H_0 . If there is 'sufficient' evidence against H_0 , one rejects it and considers the alternative hypothesis H_1 , which is the complement of H_0 , to be statistically confirmed. But standard statistical tests are never used to statistically confirm the null hypothesis. Finding no evidence against the null hypothesis H_0 does not mean that H_0 has been statistically confirmed. The reason for this imbalance is that there are two types of errors:

Error of type 1: Rejecting H_0 , although H_0 is correct.

Error of type 2: Not rejecting H_0 , although H_0 is false.

And only the probability for an error of type 1 is controlled and bounded by the test level α , with $\alpha = 0.05$ being a widely applied conventional test level. When rejecting H_0 , the test decision will be correct with at least probability $1 - \alpha$. Henceforth, when claiming that H_1 is correct, the risk of making a wrong statement is bounded by α . On the other hand, when not rejecting H_0 , it stays unclear with which probability H_0 is true. Therefore one reports that no statistical evidence against H_0 was found but without claiming that this implies that H_0 is correct.

Conceptually, every statistical test builds on a specific test statistic, whose value can be computed from the data and whose distribution under the null hypothesis is known. Because of the latter, a suitable rejection region can be specified. The rejection region is chosen such that the probability that the test statistic takes a value in this region is bounded by α if the null hypothesis H_0 is true. Thus, if H_0 is actually correct, the probability that the test statistic takes a value in the rejection region is guaranteed to be bounded by α , and so the probability for an error of type 1 will be bounded by α . To obtain a powerful test, the rejection region is chosen such that realizations within the rejection region are more likely under H_1 than under H_0 .

The p_{value} of a statistical test is the lowest test level α to which the null hypothesis could have been rejected. So to speak, the p_{value} is the probability that the test statistic takes the actually observed or an even more 'extreme' value, given that the null hypothesis H_0 is true. What is considered to be more 'extreme' depends on the test problem; see below for a concrete example.

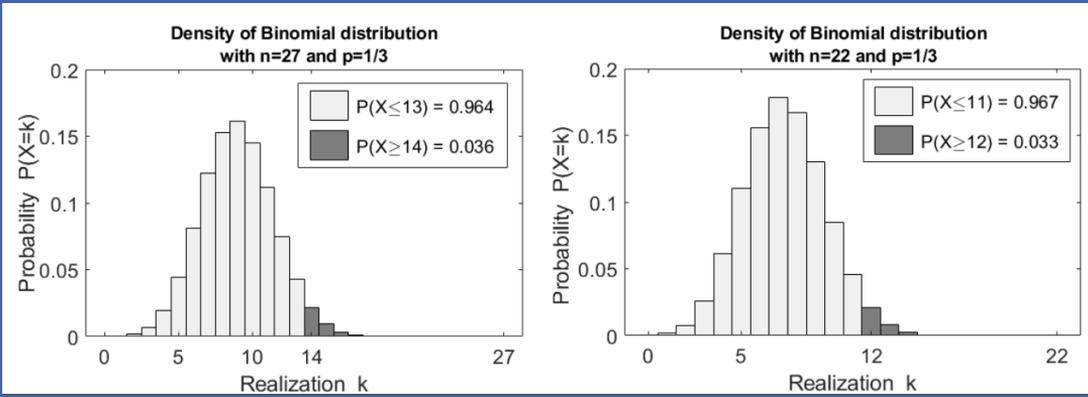


Figure 4: Densities of the Binomial distributions under H_0 . In both cases, under the hypothesis that the products do not differ by their tastes ($= H_0$), the probability for a realization within the rejection region (in dark grey) is bounded by the test level $\alpha = 0.05$.

Exact Binomial test

Under the null hypothesis H_0 that the tastes of two products do not differ, participants will nominate the correct outlier only with probability $p = \frac{1}{3}$. Mathematically, we have $H_0 : p = \frac{1}{3}$ and the number of correctly nominated outliers X will be Binomial distributed with parameters n ($n = 27$ and $n = 22$) and $p = \frac{1}{3}$. The random variable X describes the total number of correct nominations among the n participants, where under H_0 each individual participant has probability $p = \frac{1}{3}$ to nominate correctly. We note that many correct nominations support the alternative hypothesis $H_1 : p > \frac{1}{3}$, but are less likely under H_0 . The density (probability mass function) of this Binomial distribution (under $H_0 : p = \frac{1}{3}$) is

$$P(X = k) = \binom{n}{k} \cdot \left(\frac{1}{3}\right)^k \cdot \left(\frac{2}{3}\right)^{n-k} \quad (k = 0, 1, \dots, n)$$

To keep the test level below $\alpha = 0.05$, we determine the minimal critical value c that fulfills:

$$P(X \geq c) = \sum_{k=c}^n \binom{n}{k} \cdot \left(\frac{1}{3}\right)^k \cdot \left(\frac{2}{3}\right)^{n-k} \leq \alpha$$

For $n = 27$ we get $c = 14$, and for $n = 22$ we get $c = 12$. We reject H_0 in favor of the alternative hypothesis H_1 if and only if we observe realizations that are greater or equal to the critical value c . Since we have under $H_0 : P(X \geq c) \leq \alpha$, the probability for an error of type 1 is bounded by the test level α . Figure 4 shows the densities under H_0 and indicates the rejection regions for both experiments.

Given that we observe in an experiment $X=x$ correct nominations, the p_{value} of the Binomial test can be computed via:

$$p_{value} := P(X \geq x) = \sum_{k=x}^n \binom{n}{k} \cdot \left(\frac{1}{3}\right)^k \cdot \left(\frac{2}{3}\right)^{n-k}$$

That is, the p_{value} is the probability that we observe under H_0 a realization of X that is equal to or greater ($=$ 'more extreme') than the really observed value x .

Asymptotic confidence intervals

Let X be a Binomial distributed random variable with parameters n and p , and let x be the realization of X . For a sufficiently large sample size n , it follows from the asymptotic efficiency of the Maximum Likelihood estimator

$$S := \sqrt{n} \cdot \frac{(\hat{p}-p)}{\sqrt{\hat{p} \cdot (1-\hat{p})}} \sim N(0, 1), \text{ where } \hat{p} = \frac{x}{n} \text{ is the Maximum Likelihood (ML) estimator of } p.$$

That is, for large sample sizes, n , the random variable S has approximately a standard Gaussian $N(0,1)$ distribution. Henceforth, with probability $1 - \alpha$ the random variable S will take a value in between the $\frac{\alpha}{2}$ and the $(1 - \frac{\alpha}{2})$ -quantile of the $N(0,1)$ distribution, symbolically:

$$P\left(q_{\frac{\alpha}{2}} \leq S \leq q_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Solving these inequalities for p and exploiting the symmetry of the $N(0,1)$ distribution, which implies $q_{\frac{\alpha}{2}} = -q_{1-\frac{\alpha}{2}}$, one easily gets the following $1 - \alpha$ confidence interval for p

$$\hat{p} \pm q_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}},$$

where $\hat{p} = \frac{x}{n}$ is the Maximum Likelihood (ML) estimator of p and $q_{1-\frac{\alpha}{2}}$ denotes the $(1 - \frac{\alpha}{2})$ -quantile of the $N(0,1)$ distribution, so that $P(S < q_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$ if $S \sim N(0,1)$. Asymptotically, the interval covers the true unknown probability parameter p with probability $1 - \alpha$. For $\alpha = 0.05$, we have $q_{1-\frac{\alpha}{2}} = q_{0.975} = 1.96$, and we get confidence intervals with coverage probability 0.95 (=95%).

Relation between p and θ

We assume that participants can be subdivided into two disjoint groups. A first group of participants with fine sensory tastes whose group members are able to distinguish the two products, and a second group of participants whose members cannot distinguish the products and can only guess. Let θ denote the unknown fraction of people belonging to the first group. Then with probability θ a random person belongs to group 1 (i.e. can taste) and has probability 1 to identify the outlier. And with the complementary probability $1 - \theta$, the random person belongs to group 2 (i.e. cannot taste) and has only probability $\frac{1}{3}$ to guess the correct outlier. Hence, the probability that a random participant nominates the right outlier is given by:

$$p = \theta \cdot 1 + (1 - \theta) \cdot \frac{1}{3}.$$

Solving for θ yields the relationship:

$$\theta = \frac{3}{2} \cdot p - \frac{1}{2}.$$

Since this transformation between p and θ is monotone, we can transform confidence intervals for p into confidence intervals for θ by applying the transformation to the lower and upper bound of the interval.

χ^2 -test for independence in 2-by-2 tables

For lack of space, we cannot describe the χ^2 -test for independence here. Loosely speaking, the test can be used to check for statistical evidence that the probabilities for nominating the true outlier depends on factors, such as gender (male vs. female) or on whether the participant is a regular Cola drinker (Yes or No). Since we could not find any significant relationships, we omit the mathematical details behind the χ^2 -test.

Experiment	n	x	\hat{p}	Decision ($\alpha = 0.05$)	p-value
Coca vs. Pepsi	27	13	0.481	Stay with H_0	0.080
Light vs. Zero	22	12	0.545	H_1 confirmed	0.033

Table 1: Results of exact Binomial tests, where n is the number of participants. Confirming the alternative hypothesis H_1 means that the probability p for nominating the true outlier is significantly higher than $1/3$. H_1 confirmed thus means that there is statistical evidence that the two products differ in their tastes.

Experiment	\hat{p}	$\hat{\theta}$	95% CI for θ_i
Coca vs. Pepsi	0.481	0.222	[-0.060, 0.505]
Light vs. Zero	0.545	0.318	[0.006, 0.630]

Table 2: Estimators and asymptotic 95% confidence intervals (CIs) for the fractions θ of persons that can taste the difference. Negative lower bounds of CIs, like -0.060, are impossible and can be replaced by 0.

Experimental results

The results of our experiments are summarized in Tables 1-2.

Coca-Cola vs. Pepsi Cola: $x=13$ out of $n=27$ participants identified the true outlier. Although this exceeds the expected number of $e=9$ correct nominations under the null hypothesis, the result is not statistically significant to the level $\alpha = 5\%$ ($13 = x < c = 14$), so that we cannot reject the null hypothesis. Although the trend points into the direction, there is not enough empirical evidence to claim that classic Pepsi and classic Coca-Cola differ in their tastes. In particular, from the corresponding confidence interval in Table 2, it can be seen that the fraction θ of persons who can distinguish these two products might be equal to zero.

Light vs. Zero Sugar, from Coca-Cola: $x=12$ out of $n=22$ participants identified the true outlier. This also exceeds the expected number of $e=7.333$ correct nominations under the null hypothesis, and here we can reject the null hypothesis to the level $\alpha = 5\%$, since $12=x \geq c=12$. We consider this to be enough evidence to claim that the products Light and Zero Sugar (both produced by Coca-Cola) differ in their tastes. In particular, from the corresponding confidence interval in Table 2, it can be seen that the fraction θ of persons that can distinguish these two products is significantly greater than zero. However, the estimated fraction of people who can distinguish the two tastes is only equal to 0.318 (around 32%) and the low lower bound of the confidence interval indicates that it cannot be ruled out that this fraction is very low (less than 1%).

Lastly, we note that $n=19$ students participated in both triangle tests and that only $x=2$ of 19 students nominated in both cases the true outlier. The observation $x=2$ is rather close to the expected number of two correct nominations, $e = \frac{19}{9} = 2.11$, under the null hypothesis H_0 that there are no taste differences (in both cases). A Binomial test for $H_0 : p = \frac{1}{9}$ vs. $H_1 : p > \frac{1}{9}$ is far away from being significant ($p_{value} = 0.64$).

Conclusions

In both Cola experiments, it turned out to be more difficult than expected to distinguish the tastes of the two products and to identify the true outlier. We estimated the fraction of persons who can distinguish between classic Pepsi and Coca-Cola to be 22.2%, and we estimated the fraction of persons who can distinguish between Light and Zero Sugar Coca-Cola to be 31.8%. Given our rather small sample sizes, the difference between classic Pepsi and classic Coca-Cola was not statistically significant (to the test level 5%); i.e. we found no statistical evidence that the tastes of the two brands actually differ. On the other hand, there was statistical evidence that Light and Zero Sugar from Coca-Cola significantly differ in their tastes. However, surprisingly we estimated that only around one third of the drinkers can really taste this difference.

We also screened whether there are group-specific differences. Via χ^2 -test for independence we screened for significant dependencies but we could not find anything significant. The probability of identifying the true outliers seems not to depend on gender and/or whether the participant is a regular Cola drinker or not. Our final conclusion is that the taste differences between different Cola brands (Pepsi vs. Coca-Cola) and products (Light vs. Zero Sugar) seem to be much smaller than we expected and that it is potentially mainly the different marketing campaigns rather than the tastes that make people prefer the one brand over the other.

Discussions

We note that our sample sizes were rather small and that all participants were 2nd and 3rd year Bachelor students at the Faculty of Science and Engineering at Groningen University (NL). Since this implies a certain education level and a rather small age range, the participants might not be representative of typical populations. Also, our assumption that a given person can either taste the true outlier with probability one or cannot taste and has to guess what the outlier is might be over-simplistic. Another concern is that the tastes of all Cola products seem to be region-specific. That is, the Cola products of both producers might taste slightly different in different regions of the world. Our experimental results refer to Cola products that were bought in April 2022 in Dutch supermarkets; see Figures 2 and 3 for photos of the product. It might be interesting to organize another sensory experiments with larger sample sizes. Perhaps FMF might be interested in organizing a sensory experiment to compare other products, such as different beer or coffee brands.

Perio Interview: Roland van der Veen

AUTHORS: L. ÁLVAREZ HEREDA, R. MONDEN

Roland van der Veen *tied the knot* with the University of Groningen a few years ago and he is already a well-known figure in the community.

What field of mathematics are you in?

The field I work in is both topology and geometry. I like to think of myself as an all-rounder, which was my ideal as a mathematician: to understand everything.

What made you decide to study mathematics?

I thought it had some mystery to it; not necessarily school math, but 3D shapes, like cubes or icosahedrons. I can't describe why I like them, but I always did. As a kid, I would make paper airplanes and then I would end up building these shapes. These geometrical games are what sort of got me started. And you don't need to be good at solving equations to do that.

Back in high school, I wanted to become a programmer and study computer science, but that didn't happen. I think it was because my hardware was acting up so couldn't get my C++ installed properly. So I read books about it without actually programming; I ended up reading about pointers abstractly. This is how people probably read about computers in the 40s, when there were not so many available. But I think it helped me deal with all this abstract *nonsense*.

Later on, I learned that in mathematics you can still program. And now, maybe a third of the time I am either coding or debugging. Pen and paper are old-fashioned, just for show or communication; I am not one of those mathematicians who pretend that it is still the 19th century. If you want to do a computation, you do it on the computer.

And why topology, what fascinates you about it?

Knots. I like to tie knots, it's my hobby and also my way of making a living. They are a very concrete instance of topology. I like them because they are at the same time very simple and also complicated so you can attach lots of complex stuff to it, but you can also just see it, and play with it. So the instance of play, and the interaction between very abstract and then concrete aspects is what I like about this subject.

Topology in general is very abstract but also applied regarding fluid dynamics or some very, very *dirty* physics where things are way too complicated to compute but you can still have some idea of what is going on at a qualitative level; and that's how topology was born. People gave up on solving equations, but the engineers really wanted to know what was going to happen, and using topological arguments you can sometimes say something.

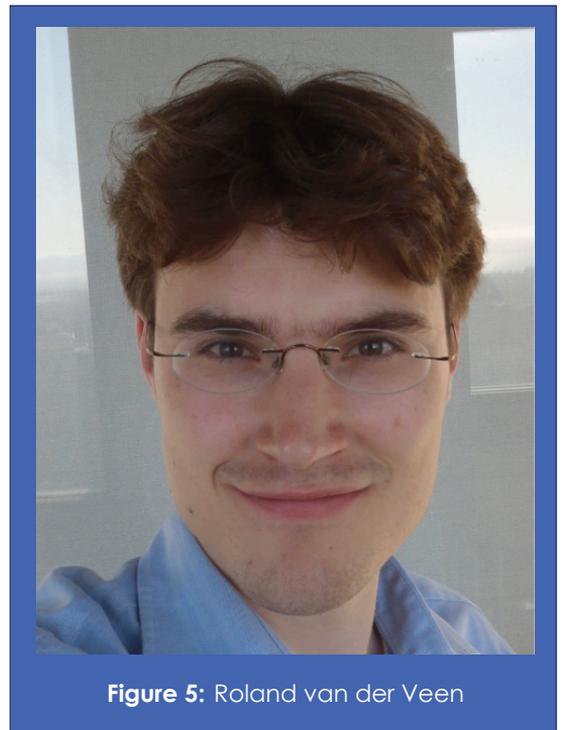


Figure 5: Roland van der Veen

If you hadn't been involved in topology, was there any other field that interested you?

I was going to do probability theory when I was doing my master's in Amsterdam. There was this very good professor who was doing probability and measure theory. Working on the underpinnings of the idea of

chance, sort of hard analysis, but also again sort of concrete intuitions about what is going to happen or not, what's likely what's not, can you turn this into a computation...? I thought that was fascinating. But things turned out differently. I definitely like doing probability stuff too, but it's a very different field, and you can tell by just opening a book, a random book, you can tell if it's probability theory or topology by looking at the type setting. Not knowing any of the symbols, but you can still see its different subject.

What part of the Netherlands do you come from originally?

Amsterdam. I was born and also attended primary school and high school there. As well as my studies and PhD; all in one place, at the UVA. Then it was time to go. But I enjoyed being there. I was also a bit arrogant, I think towards people in Groningen, or from the countryside. I had no inclination to study at another university than at the UVA. At that time I was very much under the impression that all of the other universities in Holland were sort of less. However, I was surprised that RUG was actually not the case. The level of universities is comparable, if not the same; at least in mathematics. I was just wrong, it happens sometimes.

How did you end up in Groningen?

Because I applied. I gave an FMF colloquium, that I believe was part of the application procedure but I didn't know it at the time. There was a symposium and I did my presentation with knots and they liked it. So maybe that had something to do with me working here. But otherwise, I just applied for a job. So pure luck.

"I think it helped me deal with all this abstract nonsense."

How did you deal with the change from Amsterdam to Groningen?

I lived in Rotterdam and Den Haag; my wife was working there. I was working in Leiden as well, so I have been in other places too. But with kids it is much nicer here, you can afford a bigger house, it's less crowded, less expensive and more pleasant to bike around. If you are not used to Amsterdam crowds, it can be a bit daunting. But I was very lucky to have a big place with a garden just when the pandemic hit. Before I was in a small apartment in Rotterdam, which would not have been so nice. So I was really lucky, I live in the south of Groningen, so

Helpman, where the luxury homes begin (mine is the last normal home).

What do you like best about the University of Groningen?

I arrived here before the pandemic so I still feel new. What I like best is that they gave me a job; one that is forever so I can stick around and do my own thing. Groningen is a good city and I have kids so it's about the quality of life. But, honestly, I would go anywhere to do mathematics. I was going to go to Australia or Denmark, but it didn't work out. So now I am here, and it's all good. This is kind of the life of a researcher, at least in this field. Job security is hard to come by, so if there is a permanent contract somewhere, that's what I like. Groningen is also a pleasant city.

You also applied to Denmark and Australia, so why here?

I applied to many more places, but this is where I was considering already going, and where I got the job. This is what you have to do if you want to be a researcher; you can't just stay home, because the whole world is your playing field.

Also, I did my postdocs in Berkeley for two and a half years and then came back to Holland. And I got lucky to get a grant to come back early but otherwise, I would have done a postdoc somewhere else. The postdoc life is just about traveling.

What languages do you speak, and which would you like to learn?

Just English and Dutch; some German too. I had a conversation with my colleague from Hamburg but we spoke English, and I feel embarrassed about it even though it's more convenient.

I would like to learn more languages but I just don't have the time. I am not yet done learning mathematics so since it's in the same boat as learning languages, now it's just not possible. My daughter is learning Spanish, which is cool, so I try to keep up a little bit. She's in first grade so it's just *hola* but I hope that when she progresses I will sort of hang on.

One of the nicest things about having kids is that you get to spend time with them and experience some new things that you can't otherwise. However, I am busy doing my job, it's not a 9 to 5 job, but I really work all the time.

What academic achievement are you most proud of?

I don't know. I don't think about it that way. It is not a contest or anything, and I am not doing this to be better, I just do it because I enjoy it. So maybe *proud* is not the right word to use. Mathematics is a very abstract field, and some practitioners like to make it even more abstract. Even though I study abstract mathematics, I believe it is important to keep it real and keep it in contact with the common practice; that is, keeping the users in the loop of what is going on.

In topology, people talk about knots abstractly, however, I like to relate it to concrete questions about what we know about them and what can we compute. So bringing computational techniques back into the field of knot theory is something that I am proud of. It is an ongoing process, but I think it is sort of reverting the current. Knot theory was heavily impacted by abstract physics, and many theories or proofs were given in a physical sense. Physics is an empirical science, while mathematics is not, because you have to compute and proof things in order to keep yourself honest.

I am also proud of being an all-rounder, which was always my dream.

What is your favourite equation?

What is an equation really? Because the equations that I usually write are not theory equations. Most people would say that equations are the equality of numbers and functions, but for me, that is not the case. Nonetheless, an equation that we use on a daily basis is the Kauffman bracket. Something very deep yet simple that I can explain to people with just pictures.

Any colleagues that you admire?

Mathematics, and science in general, is very fractured and split into little groups. So I don't understand what most of my colleagues do, which is very sad. There is a communication problem between scientists. Especially mathematicians, so if you ask me what my colleagues do I would go to the archive and look at past papers. But it is still difficult to assess admiration, but there are also other things like how committed they are to teaching or if they give fun talks.

The culture of giving lectures and talks, but it is not very user-friendly for students. So I admire colleagues that are able to make lectures not boring, because if they are, then who are you helping? I think there is something fundamentally wrong with the way we teach. It is not their fault, but it is a tradition of boredom.

What kind of music do you listen to?

Ah, lots of different things. I'm a really big fan of this *Tiny Desk* series of the NPR. They have this artist coming on every one or two weeks. It's always something different. It's a small setting, not like a concert and not like a recording studio, so artists have to get out of their comfort zone. They bring their instrument and maybe one or two back-ups, but in essence, it's the artists really trying to showcase. It's a very famous series, so many of the series start their performance by saying that it's a big honor to perform there and that they've watched many Tiny Desks before. The artists are all out of their comfort zone and all try really hard to play their signature songs. Often when I like something I look for what else they do, but somehow it's always less good, less fresh and less exciting than what they did on Tiny Desk.

I guess I don't have a very strong preference for a particular genre. I just like to sample. But check it out, Tiny Desk. It's really cool. Lots of coming artists were there first.

“Mathematics is a very abstract field, and some practitioners like to make it even more abstract.”

What did you have for breakfast?

Yogurt and cereal. Except I first have to help my kids. Then make sure they stop fighting. So breakfast is not an easy thing. It's not like you just go downstairs and have breakfast. No, no! Forget about that! It's not how it goes. As a parent, you have breakfast after everybody else has breakfast. Maybe. If there's still time. My sons have room next to each other and when they wake up, they go to each other's rooms. Sometimes they're just playing, but often when they haven't had breakfast yet it becomes a fight. There's also often something of a power play, because if I enter your room then I can tell you to leave. So I'm very much a parent too. I sort of gave up part of my own life, but hopefully, I will get it back soon. Like in twenty years. Well, maybe then. The last seven years were like that. You can't just go somewhere. No, no. You can't have breakfast.

Do you like to cook?

Yes, I also like to cook, I didn't tell you about that. It's something I got from my dad. I'm the cook at home. I try to learn new things. I'm not sure it's a hidden talent since I'm not sure my family would agree with that. At least I try to keep improving, that's something. I like to make Indian curries. What else? Lots of things. My kids always want pasta. I sometimes make my own pasta. Also, Indonesian cooking is quite interesting. I guess it's a Dutch thing, right? Because Indonesia was a Dutch colony, so there's a lot of heritage that sort of transformed, you know.

Is there anything else you would like to mention?

Anything that I would like to mention? I don't know, I'm surprised you didn't ask me about teaching. At least, in Dutch, the highest one can get in this university is *hoogleraar*, which literally means *high teacher*. So what I think is very important to realize is that research and teaching are not so far apart. Teaching is very important because most students will not be researchers. Most of you guys will go elsewhere and that's a good thing. So, teaching is, I think, the core business of the university, not research. Sometimes, we pretend otherwise, but I think teaching is underrated. I spend a lot of time writing new lecture notes and trying to change and improve the curriculum. But I also think about how to bridge the gap between research and teaching, because you can only teach something if you understand it. But do we really understand the basics? I guess it's rooted in how I think about music and also dance. It's very much about the fundamentals, right? The same is true in martial arts. If you can do your basic stance well, then you cannot be pushed over. Because you stand strongly. So it's not about the complications of this fancy research. That's not what it's about. It's about really understanding the basics. In teaching, it's somehow about re-examining the foundations. Even if you teach calculus for years and years, as is our sentence, I suppose, as mathematicians. Calculus is really hard and really hard to understand. It can be a really messy thing, so getting some clarity in this is very important, I think. However, this is very far from current practice.

Teaching is really important and students should also realize that student evaluations are taken very seriously. You can really change things. If many students indicate that something is not going well, then it will be discussed. If students don't say anything and are like *meh...*, then not much will change. It's like elections, right? If everybody is unhappy, but nobody votes, then nothing is going to change. If you don't like this class because the teacher was boring, then it's common not to say anything about it. However, if you speak up, you can really make a difference. On the other side, the teachers who want to change something only do so if the students also agree. If students don't say anything, then the assumption is that they are happy and therefore, things should stay as they are. This is what I think is important. To be a little bit more engaged.

I also think teaching by boredom is a problem. Teaching boring stuff. Because it really propagates. Students assume that this is how you're supposed to teach. You see it in TAs as well. There's a TA training of one day and everyone always makes fun of that. But how do the TAs learn how to teach? I guess they look at their previous TAs. They look at their teachers. But that means that if they're bad examples, then their teaching style sort of propagates. I've seen many TAs who are doing a really good job, but I've also seen many TAs who sort of assume that they should just sit there and ask the students to come to them if they have a question. Because that's what *their* TAs did. But it's really not the best way to communicate any kind of science. So I very much think this is a very important part of my job. Not just to be a researcher, but also to make teaching interesting. You could even say that I'm doing research to keep myself in shape. Just like a treadmill.

True or false?

Test your knowledge about ASML

From chipmaking to EUV and from the number of employees globally to next generation machines, discover the most important facts about our fascinating tech company.

The name 'ASML' is an acronym.

FALSE. ASML isn't an abbreviation of anything anymore, though it used to stand for 'Advanced Semiconductor Materials Lithography'. ASML was founded in 1984 as a joint venture between Philips and ASM International, so a name was chosen to reflect the partners in the venture. Over time, this name has become simply 'ASML'.

ASML makes microchips.

FALSE. ASML does not make microchips – we make the machines that other companies use to make microchips. We also don't make the silicon wafers that form the cradle of the chip. Customers such as Intel, Samsung and TSMC use ASML's DUV and EUV lithography systems to print tiny patterns on silicon that has been treated with 'photoresist' chemicals. They also rely on our metrology and inspection systems, together with our computational lithography and patterning control software solutions, to achieve the highest yield and best performance in mass production.

ASML is the only company that makes EUV (extreme ultraviolet) lithography technology.

TRUE. Unlike in the DUV (deep ultraviolet) lithography market, where ASML competes with other top-notch suppliers, ASML is currently the only lithography equipment supplier capable of producing EUV technology. Chipmakers use these EUV systems to manufacture the world's most advanced microchips. In fact, if you own a relatively new smartphone, gaming console or smart watch, chances are you've benefited directly from EUV lithography technology. We spent 20 years developing EUV with our partners and suppliers, resulting in a machine that contains around 100,000 parts. To ship just one of these huge machines to customers requires 40 freight containers, three cargo planes and 20 trucks.

An ASML machine is all you need to make microchips.

FALSE. Making chips is a complex, long and expensive process. Our customers have spent years and invested billions building 'fabs' (fabrication plants), buying equipment and training employees to become experts in the complex field of semiconductor manufacturing. ASML's lithography machines form an important part of a chipmaker's production line, but they are not all that's required to produce microchips. Lithography – printing patterns on silicon wafers – is certainly a critical step in the chipmaking process, but it's just one of many!

ASML is building a new kind of EUV lithography machine.

TRUE. In the semiconductor industry, innovation never stops. That's why we're already developing a next-generation EUV platform that increases the numerical aperture (NA) from 0.33 to 0.55. This means that the optics systems in the new machines will allow light with larger angles of incidence to hit the wafer, giving the system a higher resolution. The EUV 0.55 NA platform, called EXE, is well on its way to production – we're planning the first shipments of these machines to customers for R&D purposes by the end of 2023, and we expect them to be used in high-volume manufacturing by 2025.



At ASML, we're changemakers! Our growing team of over 37,000 people and 144 nationalities provides leading chipmakers with the hardware, software and services to mass produce patterns on silicon. We're probably part of the device you use to communicate, learn or play games with. Headquartered in Europe's prolific tech hub, the Brainport Eindhoven region in the Netherlands, we have over 60 locations globally and annual net sales of €18.6 billion in 2021. Be part of progress. Visit www.asml.com/students for more information about our events, internships, scholarships or early career opportunities.



Interested?

Contact the ASML student ambassador:

sjoerd@workingatasml.com

Humans vs. ChatGPT

A small case study at the University of Groningen

AUTHORS: M. GRZEGORCZYK AND THE HONOURS COLLEGE CONSORTIUM 2022/2023

To put into test the detectability of ChatGPT written texts, a group of UG student and staff were asked to identify whether a given paragraph about their field of study/research was written by ChatGPT or not.

Introduction

The emergence of natural language processing (NLP) technology has revolutionized the way we communicate with machines. ChatGPT, a large language model developed by OpenAI, is a prime example of this technology with its ability to generate high-quality human-like texts. However, as Artificial Intelligence (AI) language models become more sophisticated, it becomes questionable whether human readers are still able to distinguish between AI-generated and human-written texts. To shed some light on this aspect, the students of the Honours College course: 'Practical Statistical Hypothesis Testing' (taught in block 2a of the academic year 2022/2023 at the University of Groningen) designed and conducted a small case study. The aim of the study was to assess whether science staff and/or students of the University of Groningen can recognize ChatGPT-generated texts. For our study, we used ChatGPT-3, which was released on 13 February 2023. An online questionnaire with various types of texts was designed and sent to science staff and students. The questionnaire also featured demographic questions to investigate whether certain factors increase the success rate of differentiating ChatGPT-generated from human-written texts. In this paper, we give a brief overview of the outcome of the study. In forthcoming group projects, the students of the course will perform more advanced statistical analyses of the collected data. In particular, it will be checked whether the observed trends are statistically significant.

Design of the ChatGPT-3 study

Participants

For the study, we recruited 74 participants including students and staff members (professors, lecturers) of the University of Groningen. Participants were recruited through email invitations and WhatsApp messages. They were informed about the purpose of the study and received a link to the questionnaire. In particular, the institute email lists were used to invite the staff members

of the Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence to participate in our study. By coincidence, of the 74 participants, 37 were staff members and 37 were students. The group of professors consisted of 26 males, 10 females, and 1 non-binary person with an age range between 18 and 65+. The student group consisted of 21 males and 16 females, with an age range from 18 to 34. Participation was voluntary; all participants gave their consents and explicitly agreed to participate in the study prior to data collection.

Google Forms

We utilized the web-based survey tool 'Google Forms' to design the questionnaire. Google Forms enabled us to easily create a standardized questionnaire that the participants could complete online. This ensured consistency across the responses, and we could track response rates in real time. To avoid a high workload for the participants, each participant assessed only five text examples. Our original intention was to randomly select those five texts out of our pool of thirty example texts. Unfortunately, though, this could not be realized, because Google Forms does not offer any native randomization options. We overcame this by creating five different versions of the questionnaire, each containing five different randomly selected text samples. We then randomly referred each participant to one of those five questionnaires.

Text Examples

The questionnaire was structured with a few initial questions at the beginning followed by the five text examples. The questionnaire asked the participants to guess whether each text was purely human-written, or whether ChatGPT was involved in generating it. The ten pure human-written texts were sourced from existing reports originally written for courses within the Faculty of Science and Engineering (FSE) or the Bachelor Honours College Program. These ten texts were adapted to stay within a 100 word limit and checked for spelling and grammar mistakes. Subsequently, for

each human-written text, a ChatGPT counterpart was generated by asking ChatGPT-3 (released on 13 February 2023) to write an essay about the same topic. The exact prompts differed from topic to topic, but each followed a structure similar to the following phrase:

‘Write an essay introduction of approximately 100 words with the following keywords: [keywords about the topic].’

In addition to these two text types (**human-written texts** and **ChatGPT texts**), we also included texts of a third type, which we refer to as **hybrid texts**. We generated those by quickly proofreading and amending (if required) the ChatGPT texts. This was to remove standard phrases which ChatGPT uses regularly and which might make the text easier to identify; at least for people who have experience with ChatGPT. In total, our pool contained thirty text examples covering ten different topics. The students of the Honours College course provided ten human-written texts about ten different topics. For each human-written text (topic) a ChatGPT text was generated using the prompt above. Finally, slight revisions of the ten ChatGPT texts yielded the ten hybrid texts. To avoid the chance that participants encounter and cross-compare different texts about the same topic, we randomly selected the text samples with the condition that each topic could only appear once within each of the five questionnaires.

Technical details

Participants were required to have an email account associated with the University of Groningen. This was to ensure that they could only complete the questionnaire once. If the participants agreed to have their data collected, they were asked for demographic data such as their age and gender. Moreover, they were asked for their occupation — whether they were a student or a lecturer/professor, and which faculty and institute they are affiliated with. Moreover, they were asked about their experience in assessing texts written by students, their level of experience with ChatGPT, and how confident they felt in their ability to detect texts generated by ChatGPT. After the initial questions, the five text samples were provided consecutively. Each time the participant was asked to read the text and to determine whether ChatGPT was involved in generating it. In total, the questionnaire took approximately five minutes to complete.

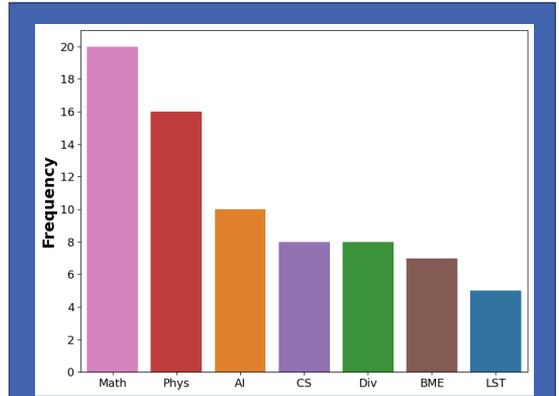


Figure 6: Number of participants per scientific field

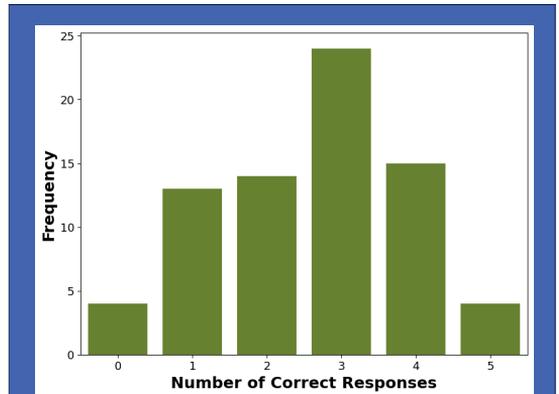


Figure 7: Marginal distribution of the scores, i.e. of the numbers of correct answers.

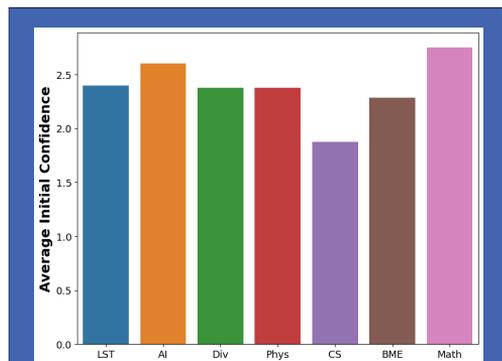
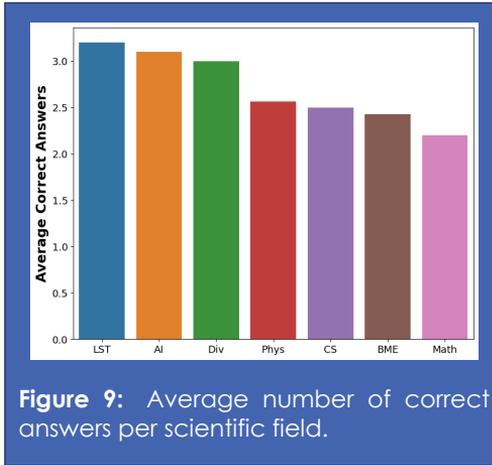


Figure 8: Average initial confidence per scientific field

Results



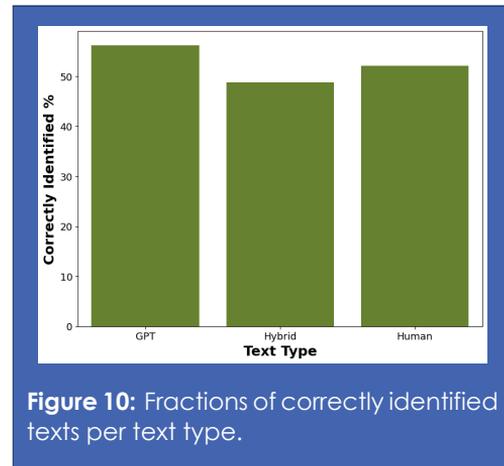
A total of 74 participants filled in the questionnaire; among them were 37 professors and 37 students. Each participant was presented with five random text samples and was asked to determine whether ChatGPT was involved in generating it. Therefore, each participant could score a minimum of 0 points and a maximum of 5 points, where each point indicates a correct answer. The expected average score if they were randomly guessing is equal to 2.5. The observed average score of the 74 participants was equal to 2.61, and so very close to 2.5. This means the average participant could not detect whether ChatGPT was used. The average scores for the students and the staff members were 2.59 and 2.62, respectively. This shows that there was no difference in the performances of staff and students.

The histogram in Figure 6 shows how the participants distribute among the different scientific fields. The participating fields were: Mathematics (Math), Physics (Phys), Artificial Intelligence (AI), Computer Science (CS), Biomedical Engineering (BME), Life sciences & Technology (LST) and Others (Div). We note that the distribution is neither representative of the entire Faculty of Science and Engineering (FSE) nor of the Bernoulli Institute (BI). Moreover, the ‘Others’ category contained members from the faculties of Economics and Business, Law, Behavioral and Social Sciences, Arts, and Spatial Sciences. The largest groups by far are Mathematics and Physics with 20 and 16 participants, respectively, while the remaining groups had only around eight participants.

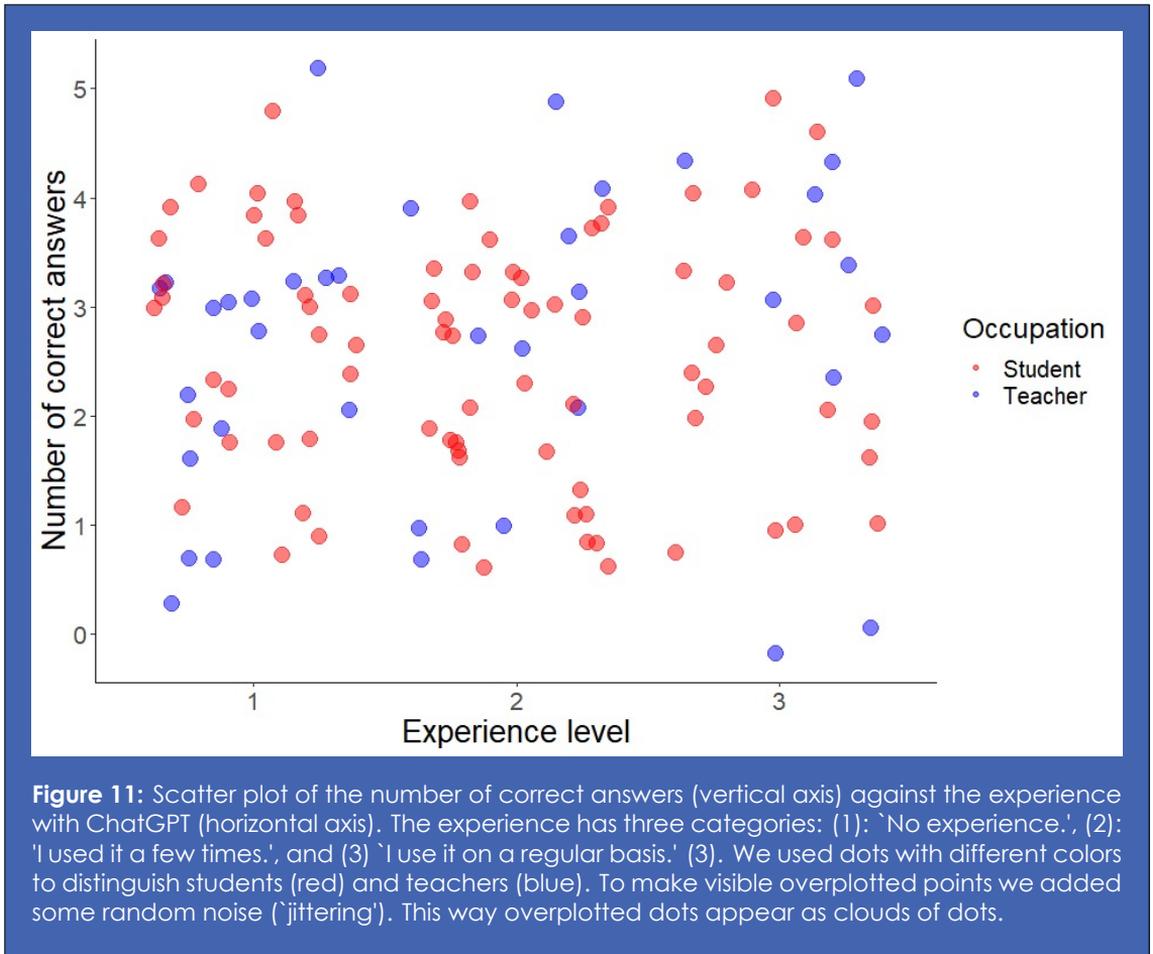
The marginal distribution of the scores (i.e. the numbers of correct answers) of all 74 participants is shown in Figure 8. Like the average score of 2.61, the marginal distribution suggests that the participants could not identify for which texts ChatGPT was used. We note that

the shape of the histogram is alike the shape of the density of a Binomial distribution with parameters $n = 74$ and $p = 0.5$. A distribution we would have expected under the assumption that all participants just randomly guess.

However, there might be performance differences between the scientific fields. To investigate this, we computed the field-specific initial confidences and the field-specific scores. The histogram in Figure 8 shows the field-specific average initial confidences. It can be seen that Mathematics (Math) and Artificial Intelligence (AI) had the highest confidences, while Computer Science (CS) had the lowest.



The average scores per field are shown in Figure 9. The group with the largest number of correct answers was Life Science and Technology (LST), with an average number of 3.2 correct responses. This was followed closely by Artificial Intelligence (AI) and Others (Div) at 3.1 and 3.0, respectively. Most remarkably, there was a surprisingly large difference between AI and Computer Sciences (CS), with AI scoring the second highest of all the groups and CS scoring the third lowest with an average of only 2.5 correct answers. This was a bit surprising, as we believed the two fields to perform very similarly. In our future work, we will employ Analysis of Variance (ANOVA) techniques to check if the field-specific differences are statistically significant. Last but not least, it seems noteworthy that Mathematics (Math) performed worst, although the Math participants started with the highest confidence.



We also computed the average fractions of correct answers per text type (ChatGPT, Hybrid and Human). The histogram in Figure 10 shows the results. All three fractions are close to 0.5 indicating that the probability for the right answer did not depend on the text type. As expected, pure ChatGPT texts were a bit easier to detect than hybrid text types. However, the difference is small and might not even be of statistical significance. In the final research step, we explored whether the scores (number of correct answers) depend on demographic factors or the experience in assessing student texts or the experience with ChatGPT. But our analyses did not reveal any clear patterns.

As an example of a typical exploratory plot, Figure 11 shows a scatter plot of the number of correct answers (vertical axis) against the experience level with ChatGPT (horizontal axis); see caption of Figure 11 for further information. In the scatter plot we do not see any trends. This suggests that regular use of ChatGPT does not increase the probability to detect whether it was used or not.

Conclusions and discussions

Most importantly, the results of our study suggest that humans cannot recognize whether ChatGPT-3 was used to generate texts (see Figure 7). There seems to be no difference between students and staff of the University of Groningen, and the performance difference between different scientific fields seems relatively small, ranging in between 2.2 and 3.2 out of five correct answers (see Figure 9). It surprised us that Mathematics was most confident but then performed worst (see Figures 8-9). We also observed that the probabilities of giving the right answer did not strongly depend on the text type. However, it seems that the use of ChatGPT was slightly easier to detect in pure ChatGPT texts than in ‘hybrid texts’, i.e. in ChatGPT texts after small revisions (see Figure 10).

Further exploratory analyses (like in Figure 11) did not reveal any trends. We note that we have presented only a first descriptive analysis of the data. More detailed statistical analyses will be performed by four student groups as part of the final assignment for the Honours College course: ‘Practical Statistical Hypothesis Testing’ in the academic year 2022/2023.

We are aware that the presented study about ChatGPT has limitations so the reported results must be interpreted with caution. Most importantly, we note that the study started with ten human-written texts and that those texts were subjectively chosen without any objective criterion. The collection of texts and the thereby covered topics can of course have had a significant influence on the results. Moreover, we note that the number of participants was relatively low and that all participants were affiliated with the University of Groningen. Among the students, the Faculty of Science and Engineering (FSE) was over-represented, while among the staff members in particular the Bernoulli Institute (which is part of FSE) was over-represented. And furthermore, among the Bernoulli Institute staff, the Mathematics department was over-represented. Also, we think that it is important to note that the participants did not get any ‘advice’ or ‘training’. We would expect that training can improve the chances of recognizing texts that were generated with ChatGPT-3.

Acknowledgements

We thank all 74 participants. The text samples are available upon request.

The Honours College Consortium 2022/2023: Frederico Remi, Jovan Andreevski, Keerthana Umesh, Mikel Martinez Garrido, Matías Sanatacruz Vallejo, Nora Totsche, Vlad Ungureanu, Hana Glumac, Anna Gumenyuk, Eline van Aalderen, Jonas Scholz, Laia Bonet Orrego, Lizaveta Yurkianets, Lana Bitar, Charlotte Hessels and J.Q Chen attended the Honours College course: ‘Practical Statistical Hypothesis Testing’ in 2022/2023. They designed the ChatGPT study, generated the questionnaires, statistically analysed the data, and wrote this paper. Marco Grzegorzczak taught the course ‘Practical Statistical Hypothesis Testing’ (PSHT) and coordinated the writing process.

Recipe

Easy Bulgarian Banitsa

AUTHOR: Y. SAVOVA

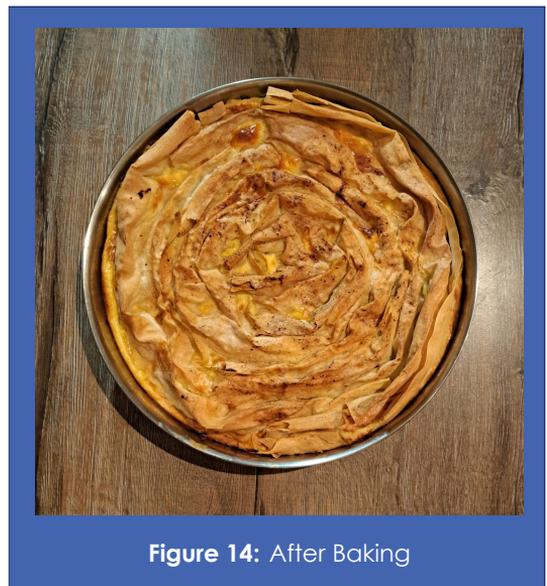
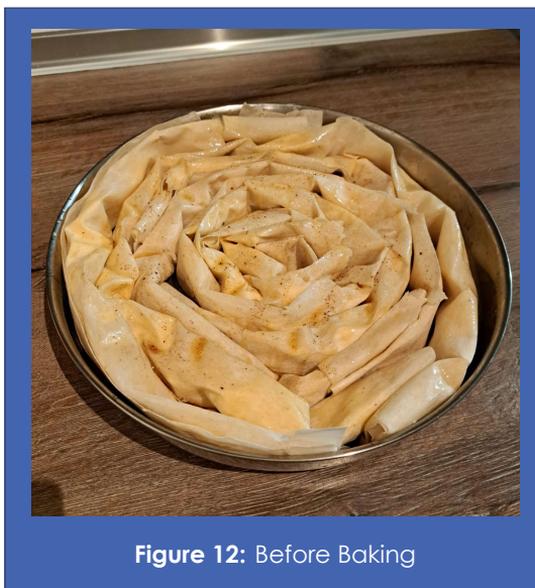
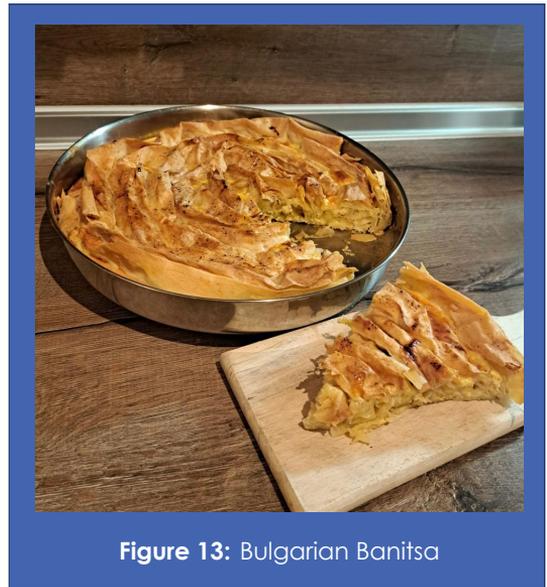
Banitsa is one of the most beloved Bulgarian dishes. It can be found on the family table on every big celebration, but it is also a quick recipe you can cook on a busy weekday. There are many version with different shapes and fillings, but here I am offering the most classic one, which is coincidentally also the easiest.

Ingredients

- 4 eggs
- 300 g white cheese (feta works)
- 60 g (4 tablespoons) yogurt
- 125 g butter
- 500 g (1 packet) pastry sheets (you can find them as “Yufka” in the Turkish supermarket Nazar)

Instructions

Preheat the oven to 180°C. Beat the eggs in a bowl. Add the yogurt and crumble the cheese in the mixture. Melt the butter in a separate bowl. Take one pastry sheet, brush it with butter and stick another one on top. Put a few tablespoons of the cheese mixture on the short side of the sheets and roll them tightly. Place the roll in the middle of a buttered baking dish and twist it into a spiral. Repeat until you have used all the cheese mixture, placing the rolls so they continue the spiral. Brush the top with the remaining melted butter. Bake it for 35-40 minutes. Let it rest for a bit and cut into triangular slices like a cake. Enjoy!



Brainwork

Fermi to the moon

AUTHORS: R. MOL

This issue's Brainwork is to solve a Fermi problem. Fermi problems are estimation problems requiring approximations and dimensional analysis to quickly estimate the solutions to large scientific calculations.

For these problems, you are not allowed to look up any values or facts such as distances, or masses. You need to rely solely on your ability to estimate orders of magnitude and the formulas and relationships that you know.

For this brainwork, we pose the following Fermi problem:

Which is greater; the gravitational force between an astronaut on the International Space Station and the Moon, or the gravitational force between the International Space Station, and all people on Earth?

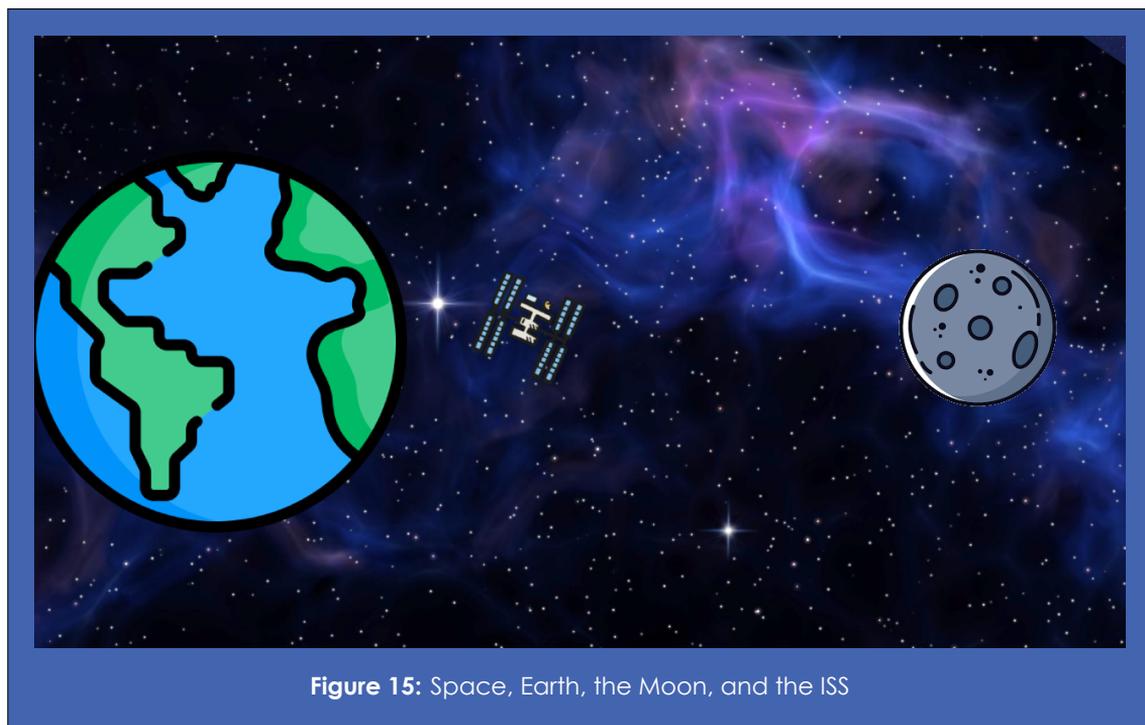
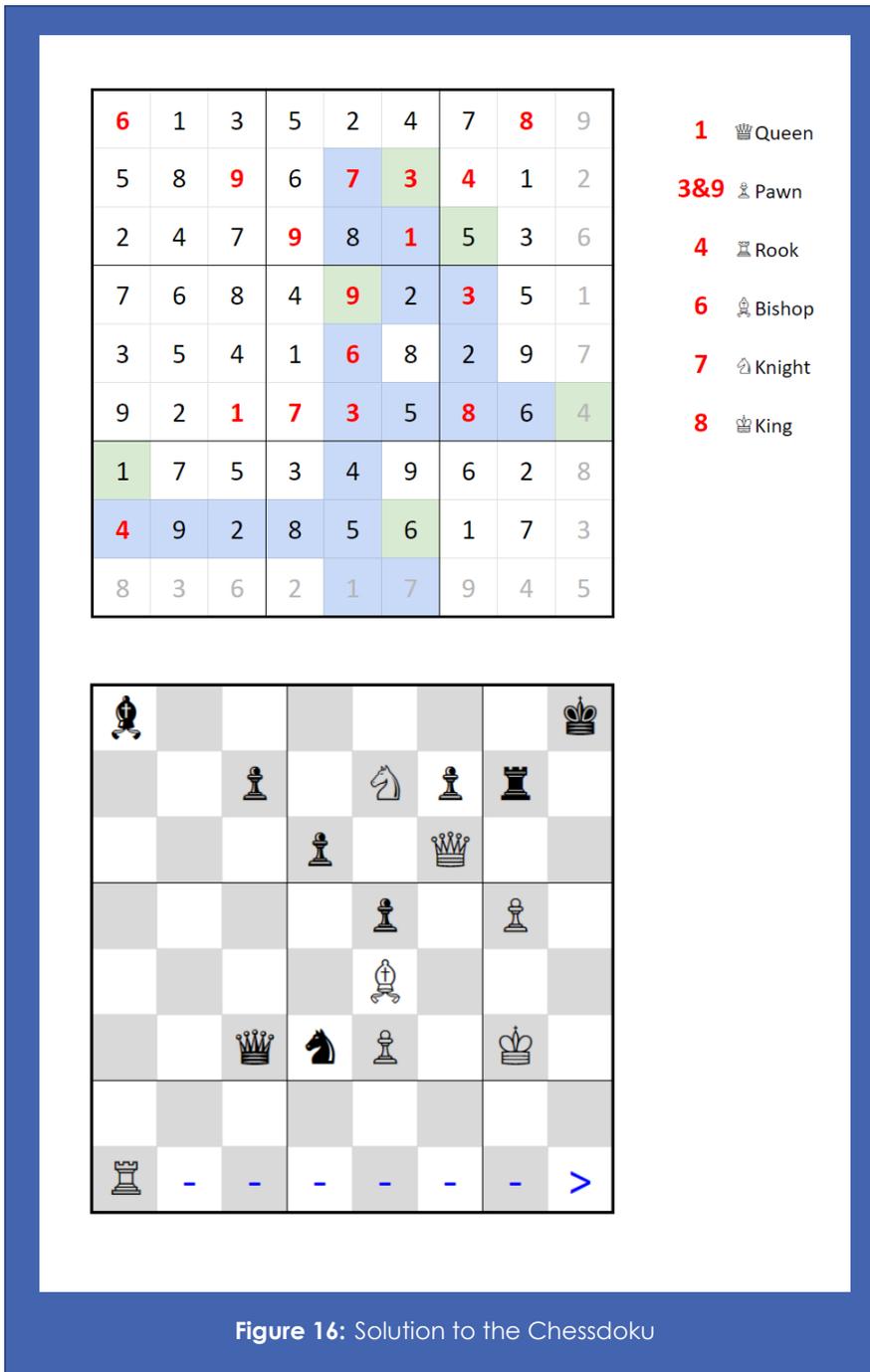


Figure 15: Space, Earth, the Moon, and the ISS

You may assume that the International Space Station is in orbit directly between the Earth and the moon.

Solution to the previous Brainwork

Below is the solution to the previous Brainwork. The numbers in red are the ones which become chess pieces, with the corresponding legend on the right. This results in the below chess board with the winning move Rh1# for white.



This puzzle was correctly solved by Jorian Pruim, Eric Jager, Egge Rouwhorst, and Armin Palavra's wife. Congratulations!

