Title :

"Now I get it !": Eureka experiences during the acquisition of mathematical concepts

Authors : Charlotte Barot[1]*, Louise Chevalier[1], Lucie Martin[1], and Véronique Izard[1]*

Affiliations :

[1] Université Paris Cité, CNRS, Integrative Neuroscience and Cognition Center, F-75006 Paris, France

* Correspondence to charlotte.barot@gmail.com or veronique.izard@u-paris.fr

Abstract

Many famous scientists have reported anecdotes where a new understanding occurred to them suddenly, in an unexpected flash. Do people generally experience such "Eureka" moments when learning science concepts? And if so, do they truly receive sudden insights when experiencing Eureka moments, or is this impression illusory? To address these questions, we developed a paradigm where participants were taught the mathematical concept of geodesic, which generalizes the common notion of straight line to straight trajectories drawn on curved surfaces. After studying lessons introducing this concept on the sphere, participants (N=56) were tested on their understanding of geodesics on the sphere and on other surfaces. Our findings indicate that Eureka experiences are common when learning mathematics, with reports by 34 (61%) participants. Moreover, Eureka experiences proved an accurate description of participants' learning, in two respects. First, Eureka experiences were associated with learning and generalization: the participants who reported experiencing Eureka's performed better at identifying counterintuitive geodesics on new surfaces. Second, and in line with the first-person experience of a sudden insight, our findings suggest that the learning mechanisms responsible for Eureka experiences are inaccessible to reflective introspection. Specifically, reports of Eureka experiences and of participants' confidence in their own understanding were associated with different profiles of performance, indicating that the mechanisms bringing about Eureka experiences and those informing reflective confidence were at least partially dissociated. Learning mathematical concepts thus appears to involve mechanisms that operate unconsciously, except when a key computational step is reached and a sudden insight breaks into consciousness.

Introduction

Learning new concepts is difficult and protracted, especially in science (Brock, 2017; Carey, 2009; Chi, 2008; diSessa, 2014; Ohlsson, 2009; Özdemir & Clark, 2007; Vosniadou, 2019; for some examples of long-term longitudinal studies see Blown & Bryce, 2006; Brock & Taber, 2020; Clark, 2006). Hence, even after several years of formal instruction, a substantial proportion of university students continue to fundamentally misunderstand key concepts from e.g. Newtonian mechanics (Caramazza, McCloskey, & Green, 1981; Clement, 1982b), biology (Dar-Nimrod & Heine, 2011; Shtulman, 2006), physics (Burgoon, Heddle, & Duran, 2011; Cohen, Eylon, & Ganiel, 1983), or mathematics (Clement, 1982a; Graeber, Tirosh, & Glover, 1989; for a recent review on persisting misconceptions in science, see Shtulman & Walker, 2020). To take an illustrative example in mathematics, many middle-schoolers fail to add, subtract, multiply or compare two fractions, or to place simple fractions on number lines (Jordan et al., 2016; Resnick et al., 2016; for similar difficulties in adults, see Post & Harel, 1991; Schneider & Siegler, 2010) – this despite the fact that fractions are typically introduced in 4th grade. Furthermore, many children, adolescents and even adults fail to apprehend that, unlike Integers, fractions form a dense set, i.e. there are infinitely many fractions between any two fractions (Smith, Solomon, & Carey, 2005; Vamvakoussi & Vosniadou, 2010; for a review see Vosniadou, Vamvakoussi, & Skopeliti, 2008). After years of study, these students thus are still at a loss with properties at the core of the concept of rational numbers.

Why is concept learning so fallible, and what happens during these long periods of time? Most authors agree that students progress towards a better understanding little by little, in a gradual manner (Carey, 2009; diSessa, 2014; Nussbaum, 1989; Ohlsson, 2009; Özdemir & Clark, 2007; Posner et al., 1982; Thornton, 1997; Vosniadou, 2002). Under this assumption, learning could be slow for several reasons: for instance, the number of incremental steps to complete could be very large, learners may often err in wrong directions instead of

progressing towards a more accurate understanding, and/or progress may be fragile. Hence, children may find support for their intuitive (and wrong) conceptions in their everyday experience (Shtulman, 2022), with experience thus acting as a counterforce constantly undoing the progress achieved in class. Alternatively, or perhaps in complement to incremental learning, learners may need to go through discrete leaps of understanding in order to acquire difficult science concepts (Clement, 1989; Gilbert & Watts, 1983) – perhaps suggestively, many theorists of conceptual change (including contenders of gradual learning) have described key learning requirements in terms of qualitative shifts (Carey, 2009; Chi, 2008; Ohlsson, 2009; Posner et al., 1982; Vosniadou et al., 2008). If this second suggestion holds, learning scientific concepts could be particularly long and fallible because the processes bringing about discrete leaps of understanding are themselves extremely fallible: they rarely complete successfully, imposing a bottleneck at key learning steps.

Are scientific concepts sometimes acquired in discrete leaps? In the making of science at least, conceptual progress sometimes *feels* discrete. For instance, the mathematician Henri Poincaré described his astounding discovery of the Fuchsian functions in these terms: "*At the moment when I put my foot on the step, the idea came to me, without anything in my former thoughts seeming to have paved the way for it (...). I did not verify the idea; I should not have had time, (...) but I felt a perfect certainty.*" (Poincaré, 1946). Like Poincaré, many scientists recounted episodes where a new understanding occurred to them suddenly, in an unexpected flash. To cite but a few, Gauss, Kekulé, and Helmholtz also reported such "Eureka experiences" leading to major advances in mathematics, chemistry, or physics (Clement, 1989; Gruber, 1995; Hadamard, 1954; Horvitz, 2002).

Eureka experiences are not reserved to privileged minds. Psychologists have identified a number of tasks, amongst them the famous Gestaltists' "insight problems", which often raise experiences similar to Poincaré's Eureka on Fuchsian functions (Ohlsson, 1984; Webb, Little,

& Cropper, 2018). Specifically, when participants solve these tasks, they produce the full solution at once rather than elaborating it progressively (Kaplan & Simon, 1990; Maier, 1931); they are not aware that they are approaching the solution, even seconds before solving the problem (Laukkonen et al., 2021; Metcalfe, 1986a, 1986b; Metcalfe & Wiebe, 1987); and moreover, the solution found is immediately perceived as correct and relevant (Danek & Wiley, 2017; Gick & Lockhart, 1995; Laukkonen et al., 2020). Importantly, and unlike the episodes recounted in scientists' memoires years after the discoveries were made, the conclusions of this research are based on measures collected while the participants were engaged in solving an experimental task, thus establishing the existence of Eureka experiences and their causal relation with the tasks at hand on a firm experimental ground.

While Eureka experiences have now been proven to arise in a range of contexts (e.g. Bowden et al., 2005; Danek et al., 2014; MacGregor & Cunningham, 2008; Webb et al., 2018), still there is little empirical work, if any, testing whether people may experience Eureka's when learning scientific concepts (for a review, see Brock, 2015). Several reports describe episodes where a student displays excitement while formulating a (correct) idea they had never expressed before, a sign that these students may have received a sudden insight (e.g. Blown & Bryce, 2006; Parnafes, 2012; Srivastava & Ramadas, 2013). Besides these case studies, to our knowledge only one study attempted to describe learning-related Eureka experiences at the population level (Liljedahl, 2005), finding reports of such experiences in 68% of students; but this study suffers from methodological limitations, casting doubts about its quantitative results. Yet, the idea that learning proceeds by sudden illuminations followed by rapid progress is common amongst professional teachers (Brock, 2015; Czarnocha & Baker, 2021) and in the general population. Many readers will probably recall episodes where they felt that they suddenly understood a notion: "Now, I get it!"

Our study was undertaken to address two main questions. First, do people generally experience Eureka moments when learning a new scientific concept? Second, if they do experience Eureka moments, is the impression to have received a sudden insight accurate or illusory? Two aspects of Eureka experiences were examined with this second question in mind. First, to assess people's impression to have gained a new understanding, we tested whether Eureka experiences are associated with genuine learning progress. Second, in a typical Eureka moment, an idea appears to break into consciousness suddenly and unexpectedly – or, to quote again Poincaré's words, this insight comes "*without anything in [one's] previous thinking seemingly paving the way for it*". To probe the veracity of this impression, we tested whether the progresses of Eureka-triggering mechanisms are accessible to reflective introspection. If learners' first-person report is accurate, these mechanisms should be unconscious, inaccessible to reflective introspection; and in particular, their progress should not inform learners' reflective judgments about their own learning.

To address these questions, we developed a paradigm where participants were taught the mathematical concept of geodesic, which generalizes the common notion of a straight line to straight trajectories drawn on curved surfaces (Spivak, 1979). Participants were given 1 to 7 lessons to learn about the geodesics of the sphere, and were then administered several tasks testing their understanding of geodesics on the sphere and on other surfaces. In addition, they were asked to reflect on their own learning and assess their confidence in their own understanding, and also to report whether they had experienced Eureka episodes during the course of the experiment.

Our analyses tested four predictions. First, before addressing questions about the relation between Eureka experiences and learning, we needed to demonstrate that our paradigm was effective in producing learning. To that aim, we checked whether participants performed better in the post-teaching tests when they had studied more lessons. Second, if concept

learning gives rise to Eureka experiences, then participants should report Eureka experiences, and these reports should be modulated by our experimental manipulation (number of lessons studied): an indication that the experiences reported are induced by the teaching phase, not by e.g. personality traits of individual participants or by the general context of the experiment. Third, if Eureka experiences reflect genuine learning progresses, participants who report Eureka's should achieve better performance in the post-teaching tests. Fourth, we compared the learning achievements associated to Eureka experiences vs. to participants' reflective judgments of confidence in their own understanding. If Eureka-triggering mechanisms are inaccessible to reflective introspection, Eureka experiences should be uniquely associated to some learning achievements, after factoring out variations in participants' judgments of confidence. Note that the reverse relation may also be true: perhaps judgments of confidence also relate to specific learning achievements, independently from the occurrence of Eureka experiences. Observing such a pattern of unique associations would indicate that the mechanisms triggering Eureka's and informing reflective confidence are at least partially dissociated, thus suggesting that the concept learning mechanisms that are responsible for Eureka experiences operate unconsciously.

Methods

Our experiment comprised four phases. In the first, *inclusion* phase, we administered two tests to select participants who had a good understanding of elementary planar geometry, but were not yet able to identify "straight lines"[1] (geodesics) on the sphere. Next, in the *teaching* phase,

---

[1] In order to help participants identify the concept presented as a generalization of planar straight lines, in our experimental material we used the common term "straight line" rather than the technical term "geodesic". Accordingly, in the description of our paradigm, we will refer to the concept under study as that of "straight line".

participants were presented with lessons about the "straight lines" of the sphere. The third phase, *test* phase, assessed participants' understanding of the concept of straight line, generalized to curved surface. Lastly, in the fourth or *Eureka* phase, participants were asked if they had experienced Eureka moments during the course of the experiment.

As explained above, in our analyses we aimed to test whether Eureka experiences were associated with genuine learning, and also whether Eureka experiences and confidence judgments were associated to the same kinds of learning. Our *test* phase was thus designed to provide a systematic assessment of participants' understanding of straight lines, with tasks probing various abilities entering into the possession of a concept. Thus, we included both identification tasks where participants needed to recognize instances of straight lines (i.e., tasks assessing the extension of their concept of straight line), as well as tasks testing their ability to draw inferences about straight lines (i.e., tasks assessing the inferential role of their concept; (for an argument that both extension and inferential role participate to the characterization of concepts, see Carey, 2009). For both identification and inference tasks, we furthermore varied the domain of application of the concept: either the sphere (the domain of application covered in the lessons), or non-sphere surfaces that were not mentioned in the lessons (on contextual effects and lack of transfer in science concept learning, see e.g. Brock & Taber, 2020). Finally, in our identification tasks we systematically crossed two variables: straightness and planarity. Indeed, based on a study testing intuitions in spherical geometry (Izard et al., 2011), we hypothesized that people would spontaneously be biased to identify straight lines with planar cuts, at least on the sphere (for direct evidence supporting this hypothesis across many surfaces, see Barot, 2022). The systematic manipulation of planarity alongside straightness thus allowed us to more finely probe the nature of the geometric properties defining "straight lines" for our participants. Specifically, if participants rely on

their spontaneous intuitions, they should respond on the basis of planarity; whereas if they apply the criteria given in the lessons, their responses should be driven by straightness.

These manipulations yielded nine different test conditions: 3 test conditions assessing participants' identification of straight lines on the sphere (non-planar straight lines do not exist on the sphere so the two variables of straightness and planarity could not be fully crossed), 4 test conditions assessing participants' identification of straight lines on various surfaces (fully crossing straightness and planarity), 1 test condition assessing inferences about straight lines on the sphere, and 1 test condition assessing inferences about straight lines on various surfaces.

Participants

Participants were recruited from a mailing list of volunteers from the greater Paris area or by word of mouth, with three inclusion criteria: being aged between 18 and 50 years, having good corrected vision, and being fluent in french. A total of 69 persons were tested, but the data of 13 participants were excluded from analyses because of their performance on inclusion tests (5 participants for poor performance in planar geometry, 3 participants for good performance in spherical geometry) and/or because of an experimenter error (6).

The final sample included 56 participants (40 females, age 18-43 years, Mean = 25.5 years, exact age missing for 10 participants)[2]. All of them had attended high school. In France, students can choose to specialize in the humanities and quit studying mathematics after

---

[2] Participants of the 1-/7- and 3-/5-lesson conditions were recruited in two successive batches. Note that the average number of lessons was the same in these two batches (4 lessons), so differences between batches could not induce biases in our analyses, where number of lessons was encoded as a numerical variable. For more information about our recruitment and testing plan, see the Supplementary Online Material.

completing 10th grade. In our sample, some participants had thus received education in mathematics only until 10th grade, while others had received up to 7 additional years of mathematics education (average number of years of education in mathematics after 10th grade: 3.9 years).

The study was conducting according to the ethical standards of Helsinki's declaration. Participants provided written informed consent before starting the experiment. The experiment lasted 60 to 90 minutes and all participants received a 15€compensation. To support people's motivation to learn, a 50€bonus was awarded to the participant who reached the highest score in each teaching condition.

Material and Procedure

Table 1 describes the content of the four phases of the experiment (inclusion phase, teaching phase, test phase, and Eureka phase)[3].

Inclusion phase

In the first phase of the experiment, participants were administered two tests assessing respectively their understanding of planar and spherical geometry. They were included in the main analyses if they had a typical (and good) understanding of elementary planar geometry but were not able to identify straight lines on spheres.

---

[3] For the sake of readability, here we only describe the tasks that are analyzed in this paper. A comprehensive list of the tasks presented to the participants can be found in the Supplementary Online Material.

Inclusion phase	Teaching phase	Test phase	Eureka phase
Planar geometry test Straight lines identification on spheres	Introduction: great circles 1, 3, 5, or 7 lessons about straight lines on the sphere	Confidence judgment (1) Straight lines identification on spheres Straight lines identification on various surfaces Confidence judgment (2) Reasoning about straight lines on the sphere and on other surfaces Confidence judgment (3)	Eureka report

Table 1. Tasks administered to the participants in the four phases of the experiment.

Planar geometry test. This test was adapted from Izard et al. (2011 planar geometry condition, questions 1-20). Participants were first introduced to a planar surface, extending indefinitely, on which points and straight lines could be drawn. Straight lines were described as lines that never turn, neither on the left nor on the right, and that continue straight ahead indefinitely. After this introduction, participants were asked a series of twenty illustrated questions about the properties of straight lines on this infinite plane. For example, in one of the trials, they were presented with a figure showing a straight line and a point and asked whether it is possible to draw a new line that goes through the point and does not intersect the first line. Questions were presented both in writing and orally through an audio recording, and participants ticked their answers (yes or no) on a response sheet. Participants were included if they made no more than 3 errors – pilot work indicated that more than 90% of geometry-educated adults should pass this criterion.

Straight lines identification on spheres. In each trial, participants were presented with a photograph of a sphere (a table tennis ball) with a line drawn on it, and were asked to indicate whether the line was "straight" or not. Three types of trials were presented (Figure 1): non-circles (non-straight; e.g. wavy line, line looping and crossing itself to form an 8 figure), great circles (straight), and small circles (non-straight, but typically judged to be straight by most adults). Each category counted 4 trials, for a total of 12 trials presented in a randomized order. Participants responded by pressing the 'O' key for yes ('oui') or the 'N' key for no ('non'). They were included if they made at least two mistakes on the small circle trials (i.e. they incorrectly judged small circles to be straight lines) or if they made at least two mistakes on the great circle trials (i.e. they incorrectly judged that great circles are not straight).

Non circle   Great circle   Small circle

Figure 1. Example stimuli from the three conditions of the spheres straight lines identification task.

Teaching phase

In the second phase, participants were taught how to generalize the concept of "straight line" to the sphere. They were first given a one-page introduction defining the "great circles" of a sphere as circles that have the same radius as the sphere on which they are drawn. This introduction also provided illustrations of great circles drawn at different orientations. Then,

participants studied 1, 3, 5, or 7 lessons about straight lines in spherical geometry. The lessons used simple physical models to explain why great circles correspond to straight lines on the sphere, but smaller circles do not. These models were: rolling a toy car on a ball (the car can be rolled along a great circle, but not along a small circle; 2 lessons), applying scotch tape on a ball (the tape wrinkles if applied along a small circle, but remains smooth if applied along a great circle; 2 lessons), pinning a rubber band on a ball (the rubber band naturally aligns with a great circle; 2 lessons), and flight routes (flight routes seem curved when mapped on a planisphere, but prove to be the shortest route and a portion of a great circle when mapped on a globe; 1 lesson). For each teaching condition, different orders of presentation were created to ensure that each lesson appeared in each position, and that a given lesson was not always followed or preceded by the same lesson.

A translated version of the teaching materials can be found on the Github repository of the project: https://github.com/charlusb/Analyses_Eurekamaths.

Test phase

At the end of the teaching phase, the first experimenter left the room and was replaced by a second experimenter, who was blind to the teaching condition assigned to the participants. This second experimenter administered three tasks: In the first task, participants judged whether lines drawn on a sphere were straight or not; in the second task they judged whether lines drawn on different surfaces were straight or not; and in the third task they were asked to draw inferences about the geometric properties of straight lines on the sphere and on other surfaces. We also measured participants' confidence in their own understanding of straight lines at different time points during the test phase.

Straight lines identification on spheres. This task was identical to the spherical geometry inclusion task.

Straight lines identification on various surfaces. Participants were presented with lines drawn on four different surfaces: cone (8 trials), cylinder (6 trials), cube (8 trials), and torus (4 trials). Each trial displayed photographs of the front and back view of a surface on which a line had been drawn (Figure 2). Participants were asked to judge whether the line presented was straight or not and indicated their answer by keypress. As explained above, we suspected that lines corresponding to the intersection of a surface with a plane would intuitively look straight, as they do on the sphere (on the sphere, planar intersections correspond to circles). Our task thus crossed the two variables of straightness (straight, not straight) and planarity (planar, non-planar): 4 trials presented non-planar non-straight lines (corresponding to non-circle lines on the sphere), 3 trials presented planar straight lines (corresponding to great circles on the sphere), 10 trials presented planar non-straight lines (corresponding to small circles on the sphere), and 9 trials presented non-planar straight lines (there are no corresponding examples on the sphere, but such lines can exist on other surfaces). Trials were presented in random order.

Reasoning about straight lines on the sphere and on other surfaces. This task consisted of a list of mathematical assertions, which participants judged to be true or false: eight assertions about the straight lines of the sphere, followed by eight assertions about straight lines on arbitrary surfaces (Table 2). The assertions were presented in a fixed order, on paper. Participants were given written definitions for the terms 'parallel' and 'perpendicular', which appeared in some of the assertions. They answered by ticking one of four response options for each assertion: 'true - certain', 'true - uncertain', 'false - uncertain', and 'false - certain'. These four options were introduced to help participants decide when they were unsure, but we only analyzed the valence of the response (true/false).
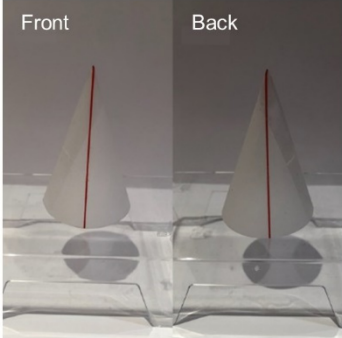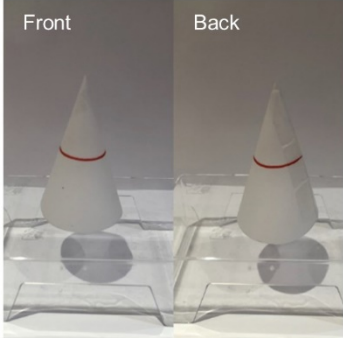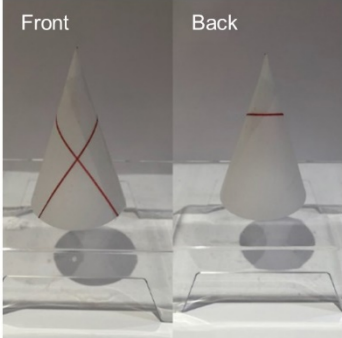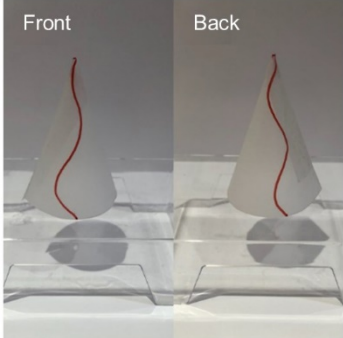
	Straight	Non-Straight
Planar	Front · Back (corresponds to great circles on the sphere)	Front · Back (corresponds to small circles on the sphere)
Non-Planar	Front · Back (does not exist on the sphere)	Front · Back (corresponds to non-circles on the sphere)

Figure 2. Example stimuli from the four conditions in the various surfaces straight lines identification task. Each trial presented two photographs showing a front and a back view of a surface and labeled as such. Three of the conditions correspond to lines that can be realized on the sphere (corresponding to great circles, small circles, and non-circles) while the last one (straight non-planar) does not exist on the sphere.

Assertion	Correct answer
1. On a sphere, given two straight lines, one can draw a straight line that intersects the first one but not the second one.	False
2. On a sphere, one can draw two straight lines that get closer to each other.	True
3. On a sphere, there is an infinity of lines perpendicular to a given line (not necessarily at the same point).	True
4. On a sphere, one can draw two straight lines that never intersect.	False
5. On a sphere, it is possible to draw two straight lines that are perpendicular.	True
6. On a sphere, two distinct straight lines always have two points of intersection.	True
7. On a sphere, it is possible to draw a straight line that is parallel to a first straight line and goes through a given point.	False
8. On a sphere, two straight lines can be drawn at a constant distance from each other.	False
9. There is a surface on which there is always one single straight line that is parallel to a first straight line and that goes through a given point.	True
10. There is a surface on which it is never possible to draw a straight line that is parallel to a first straight line and that goes through a given point.	True
11. There is a surface on which a straight line can go several times through the same point (intersecting itself).	True
12. There is a surface on which two straight lines can be drawn at a constant distance from each other.	True
13. There is a surface on which it is possible to draw several straight lines going through two given points.	True
14. There is a surface on which two straight lines are always intersecting.	True
15. There is a surface on which it is not possible to draw two perpendicular lines.	False
16. There is a surface on which it is not possible to draw two straight lines that intersect.	False

Table 2. Assertions of the reasoning task. Assertions 1-8 focus on the sphere, while assertions 9-16 invite participants to think about various surfaces. Participants were provided with the following definitions on the top of the page: "Two straight lines are parallel if they never intersect. Two straight lines are perpendicular if they intersect at a right angle". All the material presented here is translated from french.

Confidence judgments. At three different times, participants indicated how much they felt they understood the notion of straight line, on a scale graduated from 0 to 10. The first

measurement of confidence was taken at the end of the teaching phase, the second measurement after the various surfaces straight lines identification task, and the third measurement after the reasoning task. One participant in the 1-lesson group inadvertently failed to answer the third confidence question.

Eureka phase

The final part of the experiment, which was administered by the first experimenter, aimed at measuring whether participants had experienced Eureka's during the course of the experimental session. Participants were first given a description of the sensations associated with Eureka experiences (adapted from Jung-Beeman et al., 2004): episodes where a new understanding arises suddenly and unexpectedly, and is associated with a feeling of certainty. Participants indicated whether they experienced such episodes at some point during the experiment (yes or no) – this answer was used as our measure of Eureka report.

At the end of the session, participants were presented with vignettes illustrating the different phases of the experiment, so that they could indicate when exactly they had experienced Eureka's. A few participants were also asked to describe the insights that occurred to them in these occasions: they all reported insights related to the concept of straight lines on non-planar surfaces.

Analyses

We conducted analyses to address four questions: (i) whether the participants studying more lessons performed better in post-teaching tests and (ii) were more likely to report experiencing Eureka's, (iii) whether the participants reporting Eureka's showed better learning, as assessed by better performance in post-teaching tests, and (iv) whether Eureka reports and reflective judgments of confidence were associated to similar or different profiles of performance across

the post-teaching tests. Analyses were conducted in R using the packages afex and emmeans (Lenth, 2022; Singmann et al., 2021). For repeated measures analyses, we used the function mixed in package afex[4] and included a random effect of participant. For non-repeated measures analyses, we used the standard glm function. The α level was set at .05. Significant interactions involving a numerical variable (e.g. number of lessons) were explored by computing linear trends by condition (function emtrends in emmeans). Significant interactions between categorical variables were explored by computing contrasts by condition (function emmeans[5]). Holm's procedure was used to control for multiple comparisons when exploring interactions.

Analyses scripts are available on the Github repository of the project: https://github.com/charlusb/Analyses_Eurekamaths.

Effect of the number of lessons on test phase performance

First, to assess whether our paradigm effectively induced learning, we tested whether participants' objective performance in post-teaching tests varied as a function of the number of lessons studied. For the sake of simplicity, and to reduce the number of statistical tests performed, all the tasks of the test phase were analyzed in a single analysis. Accuracy was

[4] Afex' function mixed is built on glmer from lme4. To assess the significance level of each fixed effect, mixed performs a likelihood test comparing the full glmer model to restricted models obtained by setting the parameters corresponding to the effect under study to 0. In all the statistics reported here, main effects and interactions were assessed while averaging over the levels of the categorical variables of non-interest – so the results are independent of the choice of particular baseline levels for categorical variables.

[5] Function emmeans uses the full-model glmer fit from afex' mixed to compute estimated marginal means by condition. Similarly, emtrends estimates the marginal slopes associated to numerical variables. In all the statistics reported here, means and slopes were estimated while averaging over the levels of the categorical variables of non-interest. These estimated values, as well as the contrasts computed on them, are thus independent of the choice of particular baseline levels for categorical variables.

entered in a logistic mixed model analysis, with a random effect for participant, a categorical variable for test condition (total of 9 test conditions corresponding to 3 test conditions in the sphere straight line identification task: non-circle lines, great circles, small circles; 4 test conditions in the various surfaces straight line identification task: non-planar non-straight, planar straight, planar non-straight, non-planar straight; and 2 test conditions in the reasoning task: sphere, surfaces), numerical variables for teaching condition (number of lessons studied) and education in mathematics (number of years studying mathematics), as well as interactions between test condition and number of lessons, and between test condition and education in mathematics. If our manipulation was successful, we expected this analysis to yield a main effect of the number of lessons, and/or a significant interaction between test condition and number of lessons.

Effect of the number of lessons on Eureka experiences

To analyze whether studying a mathematical concept triggers Eureka experiences, we tested whether Eureka reports were modulated by the experimental manipulation introduced in the teaching phase, i.e. the number of lessons presented. Eureka reports were entered in a logistic regression with two numerical variables for the number of lessons studied and participants' years of education in mathematics. An effect of the number of lessons in this analysis would indicate that the Eureka reports observed in our study are causally related to the teaching phase, and not simply to the general context of the experiment. Furthermore, it would allow us to exclude explanations based exclusively on inter-individual variations in e.g. personality traits or education in mathematics.

Relation between Eureka experiences and test phase performance

We next tested whether the participants who experienced Eureka's reached a better level of understanding, as indicated by better performance in the test phase. To do so, we used a

logistic mixed model on accuracy with a random effect for participant and fixed effects for test condition (categorical variable with 9 levels as above), Eureka report (binary variable indicating whether the participant reported experiencing any Eureka's during the course of the experiment or not), and an interaction between test condition and Eureka report. We then conducted a second version of this analysis with additional variables for number of lessons and years of education in mathematics and their interaction with test condition. This second analysis is more conservative and detects relations between Eureka experiences and performance that cannot be explained through the influence of the teaching condition or participants' education in mathematics. However, the first analysis without covariates is potentially informative as well: if teaching condition and education in mathematics constitute the main source of variance between participants, introducing these variables as covariates may drastically reduce variability and render relations between Eureka experiences and performance impossible to detect.

Relation between Eureka experiences and confidence

The last series of analyses aimed at testing whether confidence in one's own understanding and Eureka experiences reflect similar or different learning processes. To approach this question, we first tested whether Eureka reports and confidence judgments were correlated to each other. Pairwise comparisons were conducted between four measures: the three ratings of confidence collected throughout the test phase, and the Eureka report collected at the end of the experiment. For each comparison, we conducted two Spearman correlation analyses, first without covariates, and second with number of lessons and years of education in mathematics as covariates. In each version of the analysis, *p*-values were corrected for multiple comparisons using Holm's procedure.

Second, we tested whether Eureka reports and confidence ratings were related to learning the same aspects of the target mathematical concept. To do so, we conducted a logistic mixed

model analysis on test phase accuracy with a random effect for participant and fixed effects for test condition (categorical variable with 9 levels as above), Eureka report, confidence ratings, as well as interactions of Eureka report and confidence ratings with test condition. Again, two versions of this analysis were conducted: one without variables accounting for number of lessons and years of education in mathematics, and one including these variables and their interaction with test condition as fixed effects. Since the correlation analysis found all three ratings of confidence to be highly correlated, here we used the mean of participants' three ratings. Observing significant interactions of test condition with Eureka report and confidence rating in this analysis would indicate that Eureka experiences and reflective judgments of confidence are related to different learning achievements, and would thus suggest that the processes giving rise to Eureka experiences are inaccessible to reflective introspection.

Results

Effect of the number of lessons on test phase performance

Performance varied across test conditions (main effect of test condition, $p < .001$, Table 3), ranging from 33.1% (straight line identification: straight non-planar lines on various surfaces) to 97.3% (straight line identification: great circles on spheres). In line with our expectations, the teaching phase manipulation had an impact on participants' objective performance on post-teaching tests, as attested by both a main effect of the number of lessons and a significant interaction between test condition and number of lessons. To explore the effect of the number of lessons in each test condition, we computed linear trends by number of lessons for each test condition (Figure 3). We found a positive effect of the number of lessons on participants'

ability to categorize small circles on spheres as non-straight (Estimated Trend (ET) = 0.51, 95% CI = [0.27, 0.75], $p_{corr} < .001$), as well as, more generally, on their ability to categorize planar non-straight lines on various surfaces as non-straight (same type of line as small circles on spheres; ET = 0.16, 95% CI = [0.01, 0.31], $p_{corr} = .017$). Participants who studied more lessons also performed better when asked to draw inferences about the properties of straight lines on the sphere in the reasoning test (ET = 0.18, 95% CI = [0.02, 0.35], $p_{corr} = .014$). Estimated trends were non-significant in the other test conditions (p_{corr}'s > .28; detailed results are provided in the Supplementary Online Material).

Logistic mixed model analysis of the effect of number of lessons on test phase accuracy

	df	χ^2	p
Test condition	**8**	**696.3**	**< .001**
Number of lessons	**1**	**11.9**	**< .001**
Education in Mathematics	1	2.7	.097
Test condition*Number of lessons	**8**	**36.2**	**< .001**
Test condition*Education in mathematics	8	11.0	.20

Table 3. Results of the logistic mixed model analyzing the effect of the number of lessons on test phase accuracy. Significant effects are highlighted in bold. LogLik = -1536.2, Random effect (participant): variance = .36
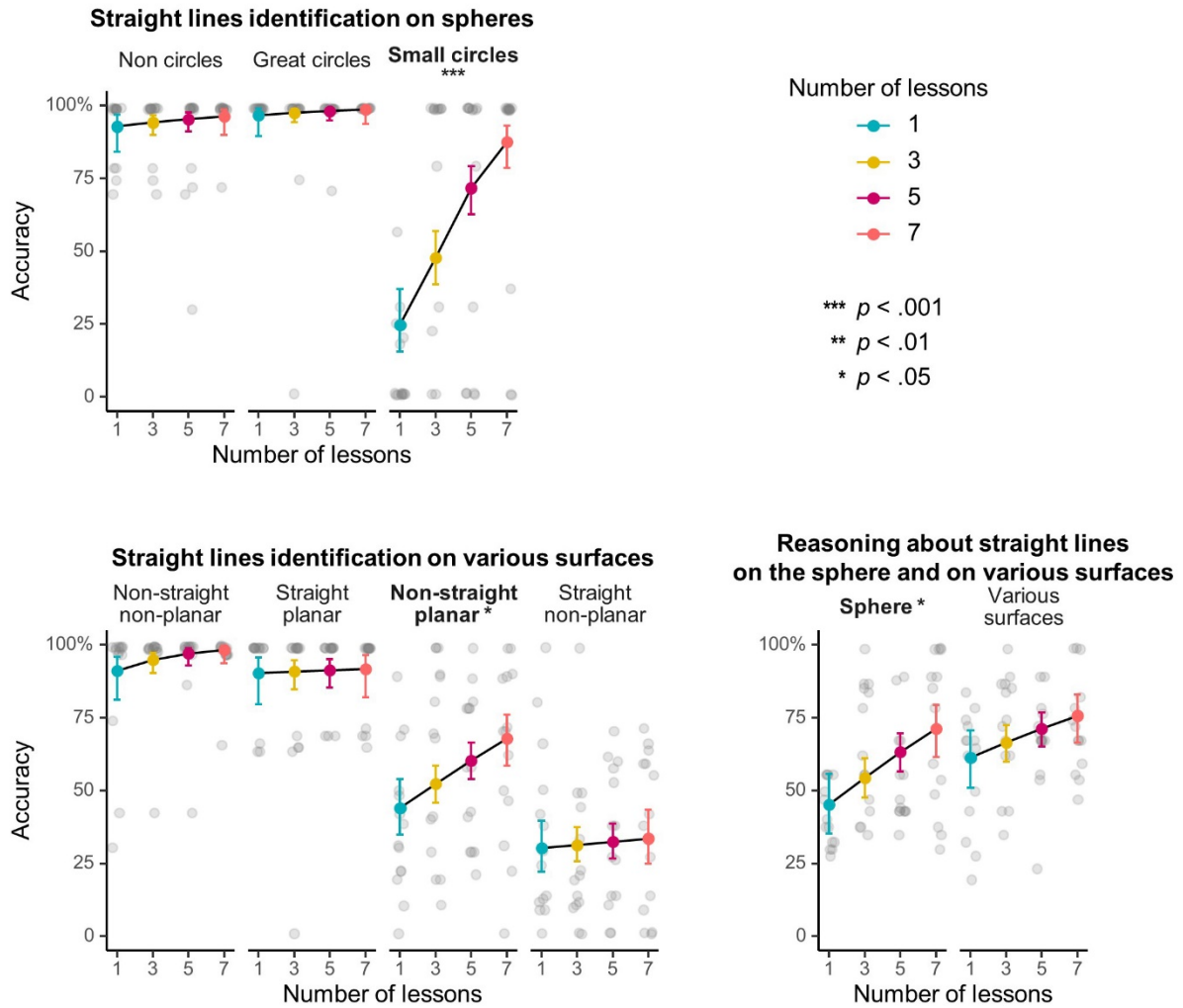
Figure 3. Effect of the number of lessons on accuracy in each test condition. Colored dots and plain lines display the predictions of the logistic mixed model. Error bars represent 95% CI. Transparent grey dots show the performance of individual participants, corrected for years of mathematics education. *P*-values from a post-hoc exploration of the linear trends by test condition, holm-corrected for multiple comparisons. Test conditions showing a significant effect of number of lessons are highlighted in bold.

Effect of the number of lessons on Eureka experiences

Thirty-four of our fifty-six participants reported experiencing a Eureka moment (61%). A logistic regression analysis (Table 4; Figure 4) revealed that Eureka reports were influenced

by the number of lessons the participants were given to read but not by their education in mathematics.

Logistic regression analysis of the effect of number of lessons on Eureka report

	χ^2	df	*p*
Number of lessons	**7.8**	**1**	**.005**
Education in mathematics	0.7	1	.42

Table 4. Results of the logistic regression analyzing the effect of the number of lessons on Eureka reports. Significant effects are highlighted in bold.
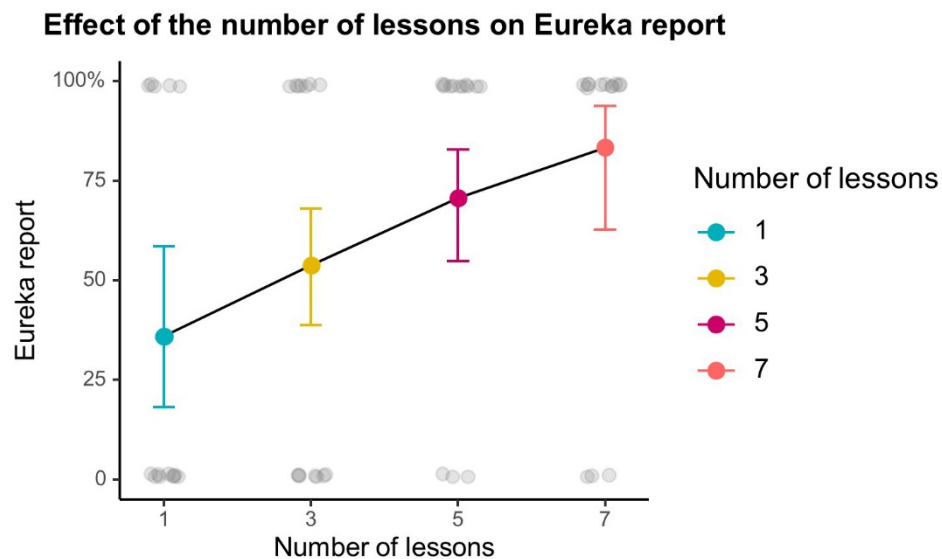
Figure 4. Effect of the number of lessons on Eureka reports. Colored dots and plain lines display the predictions of the logistic regression model. Error bars represent 95% CI. Transparent grey dots show the responses of individual participants, corrected for years of mathematics education.

Relation between Eureka experiences and test phase performance

The relation between Eureka reports and test phase performance was assessed in two mixed model analyses with and without covariates accounting for number of lessons and years of

education in mathematics. Both analyses identified a significant interaction between a variable measuring whether participants reported Eureka's and the test condition variable (Table 5), indicating that participants who did vs. did not report experiencing Eureka's differed in their performance on some of the test conditions. Exploring these interactions revealed that participants who experienced Eureka's were more likely to identify lines that are straight despite being not planar in the various surfaces straight line identification task – the type of line that does not exist on the sphere (simple model: Estimated Contrast (EC) = 1.06, 95% CI = [0.27, 1.86], p_{corr} = .002; model with covariates: EC = 1.22, 95% CI = [0.39, 2.05], p_{corr} < .001). In addition, there was a significant effect of Eureka experiences on participants' accuracy at rejecting small circles on the sphere, but only in the model that did not account for number of lessons or years of education in mathematics (simple model: EC = 0.96, 95% CI = [0.00, 1.92], p_{corr} = .043; model with covariates: EC = 0.39, 95% CI = [-0.67, 1.45], p_{corr} = 1.00). In all the other test conditions, the linear trends did not reach significance (simple model p_{corr}'s = 1.0; model with covariates: p_{corr}'s > .62; see Supplementary Online Material for detailed statistics by test condition).

Logistic mixed model analysis of the relation between Eureka report and test phase accuracy

	No covariate model			Covariates model		
	df	χ^2	*p*	df	χ^2	*p*
Test condition	**8**	**724.3**	**< .001**	**8**	**719.7**	**< .001**
Eureka report	1	0.0	.98	1	1.0	.33
Number of lessons				**1**	**12.3**	**< .001**
Education in mathematics				1	2.2	.14
Test condition*Eureka report	**8**	**27.1**	**< .001**	**8**	**34.4**	**< .001**
Test condition*Number of lessons				**8**	**43.0**	**< .001**
Test condition*Education in mathematics				8	12.0	.15

Table 5. Results of the two mixed models analyzing the relation between Eureka report and test phase accuracy. Left: simple model without covariates, right: model accounting for years of education in mathematics and number of lessons. Simple model: Loglik = -1551.5, random effect (participant): variance = 0.45. Model with covariates: LogLik -1518.5, Random effect (participant): variance = 0.36. Significant effects are highlighted in bold.

Relation between Eureka experiences and confidence

Correlation tests

Tests of the correlations between participants' three ratings of confidence and their report of Eureka experiences were conducted twice, once without covariates, and once with number of lessons and years of education in mathematics as covariates. In the two versions of the analysis, the three ratings of confidence were strongly correlated to each other (Table 6). In contrast, ratings of confidence did not correlate with Eureka reports.

Correlation table for confidence ratings and Eureka reports

	Confidence 1	Confidence 2	Confidence 3	Eureka
Confidence 1	X	**ρ(54) = .59** **_p_ < .001**	**ρ(53) = .60** **_p_ < .001**	ρ(54) = .27 _p_ = .13
Confidence 2	**ρ(52) = .57** **_p_ < .001**	X	**ρ(53) = .85** **_p_ < .001**	ρ(54) = .19 _p_ = .33
Confidence 3	**ρ(51) = .60** **_p_ < .001**	**ρ(51) = .85** **_p_ < .001**	X	ρ(53) = .17 _p_ = .33
Eureka	ρ(52) = .15 _p_ = .81	ρ(52) = .14 _p_ = .81	ρ(51) = .15 _p_ = .81	X

Table 6. Spearman's ρ coefficients and _p_-values for pairwise correlation tests. Above diagonal: without covariates, below: with number of lessons and years of education in mathematics as covariates. Significant correlations are highlighted in bold. All _p_-values were corrected for multiple comparisons using Holm's method (applied separately for the analyses with and without covariates). Note that the third rating of confidence was missing for one participant in the 1-lesson group, hence the difference in degrees of freedom. Confidence 1: measured just after participants completed the teaching phase; Confidence 2: measured after the various surfaces straight lines identification task; Confidence 3: measured after the reasoning task.

Relation between Eureka experiences, confidence and test phase performance

Lastly, we tested whether Eureka experiences and reflective judgments of confidence were related to learning the same aspects of the concept of generalized straight lines. Again, we conducted two mixed model analyses, one with and one without variables accounting for number of lessons and years of education in mathematics. The two versions of the model yielded significant interactions between Eureka report and test condition, as well as between confidence and test condition (Table 7). The model with covariates also identified a significant main effect of confidence. Exploring the interaction between Eureka and test condition revealed again a positive relation between Eureka report and performance in the condition from the various surfaces straight lines identification task that has no equivalent on the sphere (non-planar straight lines), indicating that this relation arises independently of participants' confidence in their own understanding (simple model: EC = 1.15,

95% CI = [0.32, 1.97], p_{corr} = .001; model with covariates EC = 1.26, 95% CI = [0.43, 2.10], p_{corr} < .001 .; Figure 5). None of the other test conditions showed a significant association of performance with Eureka report (simple model: p_{corr}'s > .21; model with covariates: p_{corr}'s > .68). Exploring the interaction between test condition and confidence failed to identify a specific association between confidence and performance in any of our single test conditions (simple model: p_{corr}'s > .09; model with covariates: p_{corr}'s > .29). Detailed statistics of the effects of Eureka report or confidence in each test condition are provided in the Supplementary Online Material.

Effect of Eureka experiences and confidence on the recognition of straight non-planar lines
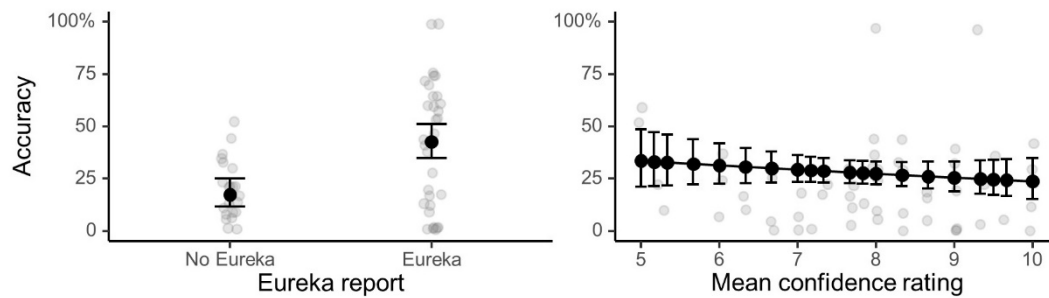

Figure 5. Relation between test phase performance, Eureka report and confidence judgments in the condition of the straight lines identification task that has no equivalent on the sphere (straight non-planar lines). A. Predicted performance for participants who did vs. did not report Eureka experiences, and individual participants' performance corrected for years of mathematics education, number of lessons, and confidence level. B. Predicted performance for participants who reported different levels of confidence, and individual participants' performance corrected for years of mathematics education, number of lessons and Eureka report. Error bars are 95% confidence intervals.

Logistic mixed model analysis of the relation between Eureka report, confidence and test phase accuracy

	No covariate model			Covariates model		
	df	χ^2	*p*	df	χ^2	*p*
Test condition	**8**	**740.5**	**< .001**	**8**	**736.8**	**< .001**
Eureka report	1	0.0	.93	1	1.4	.24
Confidence	1	1.8	.18	**1**	**4.3**	**.039**
Number of lessons				**1**	**12.0**	**< .001**
Education in mathematics				1	3.0	.085
Test condition*Eureka report	**8**	**26.1**	**.001**	**8**	**34.2**	**< .001**
Test condition*Confidence	**8**	**33.0**	**< .001**	**8**	**25.9**	**.001**
Test condition*Number of lessons				**8**	**36.1**	**< .001**
Test condition*Education in mathematics				8	10.6	.22

Table 7. Results of the two mixed models analyzing the relations between Eureka report, confidence and test phase accuracy. Left: simple model without covariates, right: model with covariates for number of lessons and education in mathematics. Simple model: LogLik = -1534.9, Random effect (participant): variance = 0.45. Covariates model: LogLik = -1505.5, Random effect (participant): variance = 0.36. Significant effects are highlighted in bold.

Discussion

Our study had four goals. First, as a prerequisite to our other aims, we needed to conceive a one-session, lab-based paradigm that effectively induces learning in mathematics. Second, using this paradigm, we asked whether learning a mathematical concept triggers Eureka experiences: episodes where a new understanding arises in a flash, suddenly and unexpectedly, accompanied with a feeling of certainty. Third, we tested whether Eureka experiences were reliable, in a first sense: that is, whether experiencing Eureka's was associated with greater learning achievements. Fourth, we probed the reliability of Eureka experiences in a second sense: in the sense that an insight is felt to break into consciousness suddenly – and thus any learning that happened beforehand must be unconscious. To get at

this question, we analyzed whether reports of Eureka experiences and reflective judgments of confidence reflected the acquisition of the same or different aspects of the target concept. If the learning mechanisms leading to sudden flashes of insight are unconscious, we reasoned, they should not contribute to inform participants' reflective judgments of confidence, and therefore some learning achievements should remain associated with Eureka experiences after factoring out variations in confidence. Below we present our findings and conclusions for each of these four questions.

Learning about geodesics

Participants were invited to study one to seven lessons introducing a mathematical concept: the concept of geodesic, generalizing the common notion of a planar straight line to straight trajectories drawn on curved surfaces. All lessons focused on the sphere and appealed to simple physical models to explain why great circles drawn on spheres correspond to lines that are straight, while small circles do not – which is counterintuitive. The models were meant to help participants ground their understanding of "straightness" on physical intuitions: for instance, a line is straight if, along this line, one can smoothly apply a band of scotch tape along it, or roll a toy car without forcing it to turn. Following this teaching phase, we assessed participants' learning in a series of tasks probing key conceptual abilities: the ability to identify straight lines, both within the domain of application covered in the lessons (on the sphere), and in new contexts (non-sphere surfaces); and the ability to draw inferences about straight lines, again manipulating the domain of application (sphere vs. various surfaces). Furthermore, our identification tasks systematically crossed line straightness against planarity (whether the line is the result of a planar cut of the surface), as this property strongly influences judgments of straightness in naïve participants (Barot, 2022). Manipulating planarity thus allowed us to assess whether participants succeeded in defining straight lines

according to the criteria presented in the lessons, even when doing so leads to a counterintuitive answer.

Our results provide evidence that our teaching phase was effective, and participants learned: reading more lessons led to better performance in several of our post-teaching test conditions. Furthermore, participants' post-test performance showed two characteristic signatures of conceptual learning. First, learning was difficult, as indicated by positive linear effects of the number of lessons on test performance in several of our test conditions. These effects show that learning was not completed after studying the first lesson, and was protracted: participants may have benefitted from the repeated presentation of the same mathematical information (great circles are straight, small circles are not straight), from the presentation of diverse physical models to ground the concept of "straightness", and/or simply from a greater incubation delay. Strikingly, these effects arose even in a test condition where participants simply needed to recall information presented in the lessons, i.e. when judging that small circles drawn on a sphere are not straight. As a second signature of conceptual learning, the content learned was inferentially rich: participants did not only memorize the information presented in the lessons but were also able to draw inferences from this information. Indeed, participants who received more lessons to study had a better performance in a test condition where they needed to generalize the notion of straight line to various surfaces, beyond the example of the sphere (condition presenting non-straight planar lines drawn on non-sphere surfaces, equivalent to small circles drawn on spheres), and in a condition where they needed to reason about the properties of straight lines on the sphere.

In our reasoning task, in particular, we found that the teaching phase enabled participants to draw non trivial inferences about the properties of straight lines in spherical geometry, and notably to realize that two straight lines drawn on a sphere can never be parallel: they necessarily cross. This property is highly counterintuitive for geometry-educated people – as

well as for people without any formal education in geometry (Izard et al., 2011). When tested on a questionnaire that inspired our planar geometry inclusion test, pretty much all U.S. adults, French teenagers, or Mundurucu people from the Amazon judged that it was possible to find parallel lines on the plane and also on the sphere (agreement on the existence of parallel lines on the sphere was above 90% in all groups). These widely shared and strongly entrenched intuitions about parallel lines resonate with the History of Mathematics. During nearly 2000 years, mathematicians attempted to demonstrate Euclid's Fifth Postulate on parallel lines from the other simpler postulates and axioms; until Gauss, Bolyai and Lobachevsky finally realized that this enterprise had started on wrong premises, and that it is possible to define geometries that are perfectly coherent with the simpler axioms but nonetheless violate the postulate on parallels (Greenberg, 2007). The geometry defined by great circles on the sphere is an example of one such coherent "non-Euclidean" geometry. In that context, it is impressive to observe that a considerable proportion of our participants questioned or even refuted the existence of parallel lines on the sphere (45.5% of negative responses for two assertions of the reasoning test claiming that parallel lines exist on the sphere). Again, the teaching phase played a role in this understanding (linear trend of number of lessons for these two assertions, ET = 0.53, 95% CI = [0.08, 0.97], p = .020, controlling for years of education in mathematics).

In other test conditions, we did not observe any impact of the teaching phase on performance. Some of these test conditions were easy and appeared compatible with people's spontaneous intuitions. In our pre-teaching inclusion test, participants generally agreed that non-circle lines drawn on a sphere are not straight (average performance 94%), or that great circles are straight (average performance 84%). Accordingly, after the teaching phase, we found that all participants reached a good performance in these two conditions, independently of the number of lessons studied (average performance: non-circle lines 94%, great circles 97%).

Performance was also good and was not affected by the number of lessons for non-planar non-straight lines (94%) and planar straight lines (90%) drawn on various surfaces – the equivalent of respectively non-circles and great circles drawn on a sphere.

In the two last test conditions, there was no effect of the teaching phase, and performance remained low in all groups, suggesting that the teaching we provided was generally insufficient to solve these tasks. These two conditions require a high level of generalization with respect to the information presented in the lessons. For example, on some surfaces it is possible to find straight lines that are not planar; however this is not possible on the sphere, and consequently, this type of line was not exemplified in the lessons. Accordingly, participants generally failed to recognize this type of straight line when tested on various surfaces (average performance 33%), and the presentation of several lessons did not seem to help. Second, the number of lessons presented had little impact on participants' reasoning about straight lines on arbitrary surfaces (average performance 67%). This last test condition included questions that required a high level of generalization (e.g. thinking about a cone to find an example of a straight line that intersects itself), as well as very intuitive questions that could be solved by thinking about the plane. Both these very easy and very hard trials presumably contributed to reducing the effects of the teaching condition in this test condition.

In summary, analyses of performance indicated that our participants benefited from the lessons presented and learned at least some aspects of the concept of generalized straight line. Our paradigm thus succeeded in creating conditions for studying the role of Eureka experiences in concept learning.

Concept learning triggers Eureka experiences

As a second conclusion, our experiment provides evidence that learning a new concept gives

rise to Eureka experiences. At the end of the session, participants were asked whether they had experienced any Eureka moments: episodes where an idea came to them suddenly and unexpectedly, accompanied with a feeling of certainty. The description provided to our participants was inspired from a study assessing the sudden insights that arise as people are tasked to solve word puzzles (Jung-Beeman et al., 2004), and covered the phenomenological components classically associated with Eureka (or "Aha!") experiences. After reading this description, a little over half of our participants reported experiencing Eureka's during the course of our experiment (61%).

Surprisingly, while the idea that students experience Eureka's in the science classroom appears to be widely shared amongst teachers (Brock, 2015), only one former study has attempted to survey Eureka experiences systematically in a population of (adult) mathematics students (Liljedahl, 2005). This study however suffers from severe methodological issues, which hampers any definitive conclusion from its results. First, reports of Eureka episodes were delayed in time until the end of a 13-week mathematics class, raising the risk of false memories. Second, and most importantly, students wrote little narratives describing their Eureka experiences in exchange of credits – and writing about a Eureka episode constituted an alternative to solving a math problem to earn these credits. Given that the population of students involved in the study (prospective teachers) reported high levels of math anxiety, this setting may have constituted a strong incentive for students to amplify or even fabricate false memories of Eureka experiences.

Importantly, in our study we incorporated several checks to ensure that Eureka reports were not purely fabricated by participants. First, we collected Eureka reports in the same session as the teaching and test phases, leaving limited time for false memories to emerge[6]. Second,

---

[6] At first view, it may have seemed tempting to instruct participants to keep an eye for Eureka episodes

participants were not rewarded for reporting Eureka's, and our instructions acknowledged that it is fully possible to learn a mathematical notion without experiencing any sudden epiphany. Third, and most importantly, we found evidence that participants' reports of Eureka episodes were modulated by our experimental manipulation, in that the groups receiving more lessons were more likely to report Eureka's. This finding attests that reports did not simply reflect variations in personality traits, or in participants' affinity with mathematics, but rather were causally related to the teaching we provided.

Fourth, Eureka experiences were related to performance in the post-teaching tests; and this relation was modulated across test conditions, in a pattern that held even when controlling for the number of lessons studied, years of education in mathematics, and participants' confidence in their own understanding. This last finding allows us to exclude several deflationary explanations. For example, it is not the case that participants confabulated and reported imaginary Eureka episodes simply because they had studied many lessons and thought they ought to understand the notion taught very well, or because they had solved the post-teaching tests easily and felt confident about their own understanding.

In summary, our findings constitute the first robust empirical evidence that learning mathematical concepts can induce Eureka experiences – or at least, experiences that are remembered as Eureka's within the short time frame of a 60- to 90-minute experimental session. Note that so far, we only discussed the *occurrence* of a certain type of phenomenological experience, leaving aside the question of the reliability of the *content* of

and report them on the spot, just as they arise. We opted against this solution, given evidence that asking people to watch their own mental processes can block performance specifically in so-called "insight problems", i.e. problems that are typically solved by sudden insight (Schooler et al. 93). We also feared that warning participants that they may experience Eureka's in the course of the experiment could lead them to expect these experiences to occur, thus increasing the rate of false positives.

these experiences. For instance, while Eureka experiences leave learners with the impression that they have just made a great leap forward in understanding, these episodes may in fact be decoupled from actual learning, arising at random times when a learner is engaged with a difficult scientific concept. Furthermore, while learners' experience is that of a sudden flash of understanding, we may find that they are able to access their own progresses at will and at all times, not only in rare Eureka episodes. These questions will be addressed in the next two sections.

Eureka experiences signal genuine learning progress

As described above, experiencing Eureka's was related to increased performance in specific subtests. In the previous section, we used this result to argue against deflationary explanations and establish a causal link between studying a concept and Eureka experiences. However, this finding also has stronger implications: it provides evidence that Eureka experiences do not simply arise when one merely engages with a novel concept, but rather signal genuine progress in learning.

According to the first-person description, learning occurs right when a Eureka is experienced: the learner was confused beforehand, and suddenly sees the light. It should be noted that, in our study, we did not attempt to assess whether Eureka experiences were accurate in their temporal aspect, and our findings are thus compatible with several alternative scenarios. For example, learning could actually be completed before the learner experiences a Eureka moment; more specifically, the knowledge acquired would remain implicit (but could still show in performance tests), until brought to the fore and made explicit in a Eureka moment (for studies testing this hypothesis in the domain of insightful problem solving, see e.g. Bowers et al., 1990; Durso, Rea, & Dayton, 1994; Ellis, Glaholt, & Reingold, 2011; Novick

& Sherman, 2003; Smith & Kounios, 1996; a critical discussion of this line of studies can be found in Ash, Cushen, & Wiley, 2009). Alternatively, it is also possible that Eureka's are felt before learning is actually completed: perhaps Eureka's are experienced during the initial steps of learning, the details of the episode being filled retrospectively once learning is complete. Again, this idea finds echoes in the literature, with theoretical views proposing that ideas gained by sudden insight are but a rough sketch or an initial hunch, further steps of evaluation and elaboration taking place after the initial discovery (e.g. Clement, 1989; Csikszentmihalyi & Sawyer, 1995; Hadamard, 1954; Ohlsson, 2009). Deciding how Eureka experiences are related in time to actual learning progress remains an interesting question for further research.

Our post-teaching tests used a variety of tasks to assess participants' understanding of the concept of generalized straight line. Interestingly, we found that Eureka experiences predicted performance in one kind of generalization test: a test probing participants' ability to identify counterintuitive straight lines of a type that had not been exemplified in the lessons (non-planar straight lines on non-sphere surfaces). To succeed in this test condition, participants needed to overcome their naïve conception associating straightness with planarity, and instead apply the defining criteria for straight lines presented in the lessons. In contrast, however, Eureka experiences were not related to performance in a task testing a different kind of generalization: drawing inferences from the information taught to reason about the properties of straight lines on the sphere and on other surfaces. This pattern suggests that Eureka experiences were triggered as participants progressed in their understanding of the definitional properties of straight lines – identifying the core properties of the concept. Eureka experiences were not triggered however when participants reflected on the consequences of adopting this new definition – deriving the inferential role of the concept of straight line. This contrast may reflect a general property of learning-related Eureka experiences: Eureka experiences would

be triggered specifically when learners engage in deep changes affecting the core definition of their concepts. Further research should be undertaken to test whether similar findings would obtain when people learn other mathematical concepts, or even concepts from other scientific domains.

Dissociations between Eureka experiences and confidence

Our experiment also aimed at evaluating another aspect of the first-person Eureka experience: the idea that some concept learning processes (the processes susceptible to trigger Eureka experiences) may operate unconsciously, and remain inaccessible to reflective introspection. To that avail, besides asking participants to report on their Eureka experiences, we also asked them to reflect on their learning and evaluate their understanding of the concept of straight line.

Reflective judgments of confidence were dissociated from reports of Eureka experiences in our findings, at two levels. First, there was no correlation between Eureka reports and confidence ratings. Second, and most importantly, Eureka reports and confidence ratings were related to different patterns of performance in the post-teaching tests. In particular, the relation between Eureka experiences and straight lines identification described above held even after factoring out variations in participants' confidence in their understanding. Learning abstract definitional properties of generalized straight lines thus involved processes that triggered Eureka experiences, yet did not inform participants' reflective evaluation of their own understanding. Conversely, we also found that confidence ratings were uniquely associated with a change in performance, independently of Eureka reports. In detail, however, while confidence was correlated to performance in one of our analysis (analysis without covariates for number of lesson and education in mathematics), none of our test conditions

showed a robust increase of performance in relation to confidence. It is thus hard to know whether the significant relation between confidence and performance reflects conceptual learning processes or processes implementing more superficial changes, such as changes in motivation or response strategies.

In summary, these findings establish the existence of dissociated processes respectively triggering Eureka experiences or informing people's reflective judgments about their own understanding. However, only the processes associated with Eureka experiences were clearly implicated in conceptual learning.

The absence of a correlation between reports of Eureka experiences and confidence ratings may seem surprising, as this finding apparently contradicts the established relation between Eureka experiences and confidence (e.g. Danek & Wiley, 2017; Laukkonen et al., 2020). Our own instructions indeed emphasized confidence (feeling "certain") as one key dimension of Eureka experiences. Participants' explanations during debriefing proved particularly enlightening to understand this unexpected finding. In particular, we observed that, in a learning context, people may very well experience sudden flashes of insight involving a sensation of certainty, without feeling any more confident *about their understanding of the concept*. For example, one of our participants described a Eureka episode where she had suddenly realized that she did not understand straight lines. This episode was experienced as a Eureka moment because the participant felt suddenly certain of her own ignorance, but she certainly did not feel any more confident about her own understanding of straight lines.

This example probably stands as an exception: the positive relation observed between Eureka reports and performance indicates that most Eureka episodes did contain information to advance people's understanding of straight lines. Eureka episodes thus probably generally led participants to feel that they had progressed in their understanding of the concept of straight line, at least in the moment. However, perhaps this sensation of certainty was transient, and

we failed to observe a correlation between Eureka experiences and confidence because we did not measure participants' confidence at the very moment where Eureka experiences occurred. Suggestively, we found that confidence was highest when measured just after the teaching phase (linear mixed model on confidence, main effect of measurement time, $F(2,105.2) = 29.9$, $p < .001$), i.e. when most Eureka experiences had just occurred (when asked to report when exactly they had experienced Eureka's, participants identified a total of 65 episodes, 44 of which were situated during the inclusion or teaching phase) and participants had not yet confronted their understanding to the generalization tests. Knowledge that has been gained by sudden insight may be particularly susceptible to interference from further testing: just like people solving problems by insights cannot describe how they found the solution (Bowden, 1997; Maier, 1931; Schooler & Melcher, 1995), in learning contexts Eureka experiences may convey knowledge without conscious access to any epistemic justification. Hence, while learners may be able to generalize the knowledge they gain by insight to new situations, doing so may lead them to realize that they cannot explain why these inferences are founded, and thus to lose confidence in their understanding.

In terms of mechanisms, our findings suggest that conceptual learning may involve an interplay between progressive learning processes operating outside the scope of consciousness, and consciousness acting as a discrete filter for access to learned information (for a model presenting consciousness as a discrete filter on perceived information, see Dehaene, 2014; Dehaene, Sergent, & Changeux, 2003; for a similar idea applied to insightful problem solving, see Bowers et al., 1990; Ellis et al., 2011; Jung-Beeman et al., 2004; Zhong, Dijksterhuis, & Galinsky, 2008). Conceptual learning has been modelled computationally as a progressive accumulation process (e.g. Bonawitz et al., 2019; Gopnik & Wellman, 2012), where the learner would constantly evaluate how her conceptual representations fit with newly received information, and engage in conceptual change when a competing

representation overcomes the current one. Eureka experiences could reflect key computational steps in these models: they could be triggered for example when a competing representation reaches a certain threshold (Eureka about a new understanding), or when the current representation drops to a floor level (Eureka about one's ignorance).

Under this view, while learning may progress incrementally at the unconscious level, overt changes in concepts would necessarily involve the experience of a sudden leap of understanding. Interestingly, performance in our most difficult test of straight line identification (non-planar straight lines on non-sphere surfaces) appears coherent with this suggestion: participants did not succeed in this condition unless they had experienced a Eureka moment (see Figure 5). We suggest that the particular difficulty of concept learning may reside in this interplay between unconscious and conscious mechanisms. Hence, learning mechanisms may rarely reach the stage where progresses are made accessible to consciousness, such that students appear to stagnate for long periods of time. Moreover, after a Eureka experience has broadcasted a new leap in understanding, unless this insight is supported by external feedback it would remain isolated, not grounded in epistemic justifications, and thus the progress achieved would be particularly fragile.

Conclusion

We developed a paradigm where participants were taught and tested on a novel mathematical concept, and analyzed participants' reports of Eureka experiences in regard to their learning performance. More than half of our participants reported experiencing Eureka moments during the experimental session; and these episodes signaled genuine learning achievements, in the sense that participants who experienced Eureka's hold a more accurate and more generalizable representation of the concept taught than those who did not experience

Eureka's. Moreover, the progresses associated with Eureka experiences failed to inform participants' reflective judgments about their own learning. Our findings thus provide evidence that the Eureka experiences that arise while learning mathematical concepts are accurate, in two senses: they reflect the functioning of concept learning processes, which appear to be inaccessible to reflective introspection.

Our study constitutes a first step in the investigation of learning-related Eureka experiences, and thus raises several questions for future research. First, are the Eureka experiences observed in the contexts of concept learning and problem solving qualitatively different, or do they reflect similar psychological processes? Second, if all Eureka experiences turn out to indicate the successful termination of a search process, what is the nature of the search involved in conceptual learning, when all the necessary information has been provided explicitly by a teacher? Third, what kinds of learning achievements are related to Eureka experiences, and could a better understanding of the nature of these experiences and the conditions under which they arise help design more effective teaching methods for science and mathematics? We hope that our research will spark interest in these questions.

Acknowledgements

References

Ash, I. K., Cushen, P. J., & Wiley, J. (2009). Obstacles in investigating the role of restructuring in insightful problem solving. *The Journal of Problem Solving, 2*(2), 6-41. https://doi.org/10.7771/1932-6246.1056

Barot, C. (2022). *Concept learning in mathematics: From straight lines to geodesics.* (PhD), Sorbonne Université.

Blown, E. J., & Bryce, T. G. K. (2006). Knowledge Restructuring in the Development of Children's Cosmologies. *International Journal of Science Education, 28*(12), 1411-1462. https://doi.org/10.1080/09500690600718062

Bonawitz, E., Ullman, T. D., Bridgers, S., Gopnik, A., & Tenenbaum, J. B. (2019). Sticking to the Evidence? A Behavioral and Computational Case Study of Micro-Theory Change in the Domain of Magnetism. *Cognitive Science, 43*(8), e12765. https://doi.org/10.1111/cogs.12765

Bowden, E. M. (1997). The Effect of Reportable and Unreportable Hints on Anagram Solution and the Aha! Experience. *Consciousness and Cognition, 6*(4), 545-573. https://doi.org/https://doi.org/10.1006/ccog.1997.0325

Bowden, E. M., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight. *Trends in Cognitive Sciences, 9*(7), 322-328. https://doi.org/10.1016/j.tics.2005.05.012

Bowers, K. S., Regehr, G., Balthazard, C., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive Psychology, 22*(1), 72-110. https://doi.org/10.1016/0010-0285(90)90004-N

Brock, R. (2015). Intuition and insight: two concepts that illuminate the tacit in science education. *Studies in Science Education, 51*(2), 127-167. https://doi.org/10.1080/03057267.2015.1049843

Brock, R., & Taber, K. S. (2020). Making claims about learning: a microgenetic multiple case

study of temporal patterns of conceptual change in learners' activation of force conceptions. *International Journal of Science Education, 42*(8), 1388-1407. https://doi.org/10.1080/09500693.2020.1764657

Brock, R. A. (2017). *Making sense of making sense: A microgenetic multiple case study of five students' developing conceptual compounds related to physics.* (Doctoral thesis), University of Cambridge.

Burgoon, J. N., Heddle, M. L., & Duran, E. (2011). Re-Examining the Similarities Between Teacher and Student Conceptions About Physical Science. *Journal of Science Teacher Education, 22*(2), 101-114. https://doi.org/10.1007/s10972-010-9196-x

Caramazza, A., McCloskey, M., & Green, B. (1981). Naive beliefs in "sophisticated" subjects: misconceptions about trajectories of objects. *Cognition, 9*(2), 117-123. https://doi.org/10.1016/0010-0277(81)90007-X

Carey, S. (2009). *The Origin of Concepts*. New York, NY: Oxford University Press.

Chi, M. T. H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *Handbook of research on conceptual change* (pp. 61-82). New York: Routledge.

Clark, D. B. (2006). Longitudinal Conceptual Change in Students' Understanding of Thermal Equilibrium: An Examination of the Process of Conceptual Restructuring. *Cognition and Instruction, 24*(4), 467-563. https://doi.org/10.1207/s1532690xci2404_3

Clement, J. (1982a). Algebra Word Problem Solutions: Thought Processes Underlying a Common Misconception. *Journal for Research in Mathematics Education, 13*(1), 16-30. https://doi.org/10.5951/jresematheduc.13.1.0016

Clement, J. (1982b). Students' preconceptions in introductory mechanics. *American Journal of Physics, 50*(1), 66-71. https://doi.org/s10.1119/1.12989

Clement, J. (1989). Learning via Model Construction and Criticism. In J. A. Glover, R. R.

Ronning, & C. R. Reynolds (Eds.), *Handbook of Creativity* (pp. 341-381). Boston, MA: Springer US.

Cohen, R., Eylon, B., & Ganiel, U. (1983). Potential difference and current in simple electric circuits: A study of students' concepts. *American Journal of Physics, 51*(5), 407-412. https://doi.org/10.1119/1.13226

Csikszentmihalyi, M., & Sawyer, K. (1995). Creative Insight: The Social Dimension of a Solitary Moment. In J. E. Davidson & R. J. Sternberg (Eds.), *The Nature of Insight* (pp. 329-363). Cambridge, MA: The MIT Press.

Czarnocha, B., & Baker, W. (Eds.). (2021). *Creativity of an Aha! Moment and Mathematics Education*. Leiden: Brill.

Danek, A. H., Fraps, T., von Müller, A., Grothe, B., & Öllinger, M. (2014). Working Wonders? Investigating insight with magic tricks. *Cognition, 130*(2), 174-185. https://doi.org/10.1016/j.cognition.2013.11.003

Danek, A. H., & Wiley, J. (2017). What about False Insights? Deconstructing the Aha! Experience along Its Multiple Dimensions for Correct and Incorrect Solutions Separately. *Frontiers in Psychology, 7*(2077). https://doi.org/10.3389/fpsyg.2016.02077

Dar-Nimrod, I., & Heine, S. J. (2011). Genetic essentialism: on the deceptive determinism of DNA. *Psychological bulletin, 137*(5), 800-818. https://doi.org/10.1037/a0021860

Dehaene, S. (2014). *Consciousness and the brain: deciphering how the brain codes our thoughts*. New York, NY: Penguin Books.

Dehaene, S., Sergent, C., & Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences, 100*(14), 8520-8525. https://doi.org/10.1073/pnas.1332574100

diSessa, A. A. (2014). A History of Conceptual Change Research: Threads and Fault Lines. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2 ed., pp. 88-108). Cambridge: Cambridge University Press.

Durso, F. T., Rea, C. B., & Dayton, T. (1994). Graph-Theoretic Confirmation of Restructuring during Insight. *Psychological Science, 5*(2), 94-98. https://doi.org/10.1111/j.1467-9280.1994.tb00637.x

Ellis, J. J., Glaholt, M. G., & Reingold, E. M. (2011). Eye movements reveal solution knowledge prior to insight. *Consciousness and Cognition, 20*(3), 768-776. https://doi.org/10.1016/j.concog.2010.12.007

Gick, M. L., & Lockhart, R. S. (1995). Cognitive and Affective Components of Insight. In J. E. Davidson & R. J. Sternberg (Eds.), *The Nature of Insight* (pp. 197-228). Cambridge, MA: The MIT Press.

Gilbert, J. K., & Watts, D. M. (1983). Concepts, Misconceptions and Alternative Conceptions: Changing Perspectives in Science Education. *Studies in Science Education, 10*(1), 61-98. https://doi.org/10.1080/03057268308559905

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. *Psychological bulletin, 138*(6), 1085-1108. https://doi.org/10.1037/a0028044

Graeber, A. O., Tirosh, D., & Glover, R. (1989). Preservice Teachers' Misconceptions in Solving Verbal Problems in Multiplication and Division. *Journal for Research in Mathematics Education, 20*(1), 95-102. https://doi.org/10.2307/749100

Greenberg, J. (2007). *Euclidean and Non-Euclidean Geometries: Development and History*. W. H. Freeman.

Gruber, H. E. (1995). Insight and affect in the history of science. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight.* (pp. 397-431). Cambridge, MA: The MIT

Press.

Hadamard, J. (1954). *The Psychology of Invention in the Mathematical Field*. New York, NY: Dover Publications.

Horvitz, L. A. (2002). *Eureka! Scientific Breakthroughs that Changed the World*. New York, NY: John Wiley & Sons.

Izard, V., Pica, P., Spelke, E. S., & Dehaene, S. (2011). Flexible intuitions of Euclidean geometry in an Amazonian indigene group. *Proceedings of the National Academy of Sciences, 108*(24), 9782-9787. https://doi.org/10.1073/pnas.1016686108

Jordan, N. C., Resnick, I., Rodrigues, J., Hansen, N., & Dyson, N. (2016). Delaware Longitudinal Study of Fraction Learning: Implications for Helping Children With Mathematics Difficulties. *Journal of Learning Disabilities, 50*(6), 621-630. https://doi.org/10.1177/0022219416662033

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., Reber, P. J., & Kounios, J. (2004). Neural Activity When People Solve Verbal Problems with Insight. *PLoS Biology, 2*(4), e97. https://doi.org/10.1371/journal.pbio.0020097

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22*(3), 374-419. https://doi.org/10.1016/0010-0285(90)90008-R

Laukkonen, R. E., Ingledew, D. J., Grimmer, H. J., Schooler, J. W., & Tangen, J. M. (2021). Getting a grip on insight: real-time and embodied Aha experiences predict correct solutions. *Cognition and Emotion, 35*(5), 918-935. https://doi.org/10.1080/02699931.2021.1908230

Laukkonen, R. E., Kaveladze, B. T., Tangen, J. M., & Schooler, J. W. (2020). The dark side of Eureka: Artificially induced Aha moments make facts feel true. *Cognition, 196*, 104122. https://doi.org/10.1016/j.cognition.2019.104122

Lenth, R. V. (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means. R
package version 1.7.2. https://CRAN.R-project.org/package=emmeans

Liljedahl, P. G. (2005). Mathematical discovery and affect: the effect of AHA! experiences on
undergraduate mathematics students. *International Journal of Mathematical
Education in Science and Technology, 36*(2-3), 219-234.
https://doi.org/10.1080/00207390412331316997

MacGregor, J. N., & Cunningham, J. B. (2008). Rebus puzzles as insight problems. *Behav
Res Methods, 40*(1), 263-268. https://doi.org/10.3758/BRM.40.1.263

Maier, N. R. F. (1931). Reasoning in humans. II. The solution of a problem and its appearance
in consciousness. *Journal of Comparative Psychology, 12*, 181-194.
https://doi.org/10.1037/h0071361

Metcalfe, J. (1986a). Feeling of knowing in memory and problem solving. *Journal of
Experimental Psychology: Learning, Memory, and Cognition, 12*, 288-294.
https://doi.org/10.1037/0278-7393.12.2.288

Metcalfe, J. (1986b). Premonitions of insight predict impending error. *Journal of
Experimental Psychology: Learning, Memory, and Cognition, 12*(4), 623-634.
https://doi.org/10.1037/0278-7393.12.4.623

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving.
Memory & Cognition, 15(3), 238-246. https://doi.org/10.3758/BF03197722

Novick, L. R., & Sherman, S. J. (2003). On the nature of insight solutions: Evidence from
skill differences in anagram solution. *The Quarterly Journal of Experimental
Psychology Section A, 56*(2), 351-382. https://doi.org/10.1080/02724980244000288

Nussbaum, J. (1989). Classroom conceptual change: philosophical perspectives. *International
Journal of Science Education, 11*(5), 530-540.
https://doi.org/10.1080/0950069890110505

Ohlsson, S. (1984). Restructuring revisited. *Scandinavian Journal of Psychology, 25*(1), 65-78. https://doi.org/10.1111/j.1467-9450.1984.tb01001.x

Ohlsson, S. (2009). Resubsumption: A Possible Mechanism for Conceptual Change and Belief Revision. *Educational Psychologist, 44*(1), 20-40. https://doi.org/10.1080/00461520802616267

Özdemir, G., & Clark, D. B. (2007). An overview of conceptual change theories. *Eurasia Journal of Mathematics, Science and Technology Education, 3*(4), 351-361. https://doi.org/10.12973/ejmste/75414

Parnafes, O. (2012). Developing Explanations and Developing Understanding: Students Explain the Phases of the Moon Using Visual Representations. *Cognition and Instruction, 30*(4), 359-403. https://doi.org/10.1080/07370008.2012.716885

Poincaré, H. (1946). *The Foundations of Science: Science and Hypothesis, the Value of Science, Science and Method* (G. B. Halsted, Trans.). Lancaster, PA: The Science Press.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science education, 66*(2), 211-227.

Post, T. R., & Harel, G. (1991). Intermediate teachers' knowledge of rational number concepts. In E. Fennema, T. P. Carpenter, & S. J. Lamon (Eds.), *Integrating research on teaching and learning mathematics* (pp. 177-198). Albany: State University of New York Press.

Resnick, I., Jordan, N. C., Hansen, N., Rajan, V., Rodrigues, J., Siegler, R. S., & Fuchs, L. S. (2016). Developmental growth trajectories in understanding of fraction magnitude from fourth through sixth grade. *Developmental Psychology, 52*, 746-757. https://doi.org/10.1037/dev0000102

Schneider, M., & Siegler, R. S. (2010). Representations of the magnitudes of fractions. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1227-1238. https://doi.org/10.1037/a0018170

Schooler, J. W., & Melcher, J. (1995). The Ineffability of Insight. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The Creative Cognition Approach* (pp. 97-133). Cambridge, MA: The MIT Press.

Shtulman, A. (2006). Qualitative differences between naïve and scientific theories of evolution. *Cognitive Psychology, 52*(2), 170-194. https://doi.org/10.1016/j.cogpsych.2005.10.001

Shtulman, A. (2022). Navigating the conflict between science and intuition. In M. Bélanger, P. Potvin, S. Horst, A. Shtulman, & E. F. Mortimer (Eds.), *Multidisciplinary Perspectives on Representational Pluralism in Human Cognition* (pp. 122-140). New York: Routledge.

Shtulman, A., & Walker, C. (2020). Developing an Understanding of Science. *Annual Review of Developmental Psychology, 2*(1), 111-132. https://doi.org/10.1146/annurev-devpsych-060320-092346

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). afex: Analysis of Factorial Experiments. R package version 1.0-1. https://CRAN.R-project.org/package=afex

Smith, C. L., Solomon, G. E., & Carey, S. (2005). Never getting to zero: Elementary school students' understanding of the infinite divisibility of number and matter. *Cognitive Psychology, 51*(2), 101-140. https://doi.org/10.1016/j.cogpsych.2005.03.001

Smith, R. W., & Kounios, J. (1996). Sudden insight: All-or-none processing revealed by speed–accuracy decomposition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1443-1462. https://doi.org/10.1037/0278-7393.22.6.1443

Spivak, M. D. (1979). *A comprehensive introduction to differential geometry*. Berkeley, CA: Publish or Perish.

Srivastava, A., & Ramadas, J. (2013). Analogy and Gesture for Mental Visualization of DNA Structure. In D. F. Treagust & C.-Y. Tsui (Eds.), *Multiple Representations in Biological Education* (pp. 311-329). Dordrecht: Springer Netherlands.

Thornton, R. K. (1997). Conceptual dynamics: Following changing student views of force and motion. *AIP Conference Proceedings, 399*(1), 241-266. https://doi.org/10.1063/1.53137

Vamvakoussi, X., & Vosniadou, S. (2010). How Many Decimals Are There Between Two Fractions? Aspects of Secondary School Students' Understanding of Rational Numbers and Their Notation. *Cognition and Instruction, 28*(2), 181-209. https://doi.org/10.1080/07370001003676603

Vosniadou, S. (2002). On the Nature of Naïve Physics. In M. Limón & L. Mason (Eds.), *Reconsidering Conceptual Change: Issues in Theory and Practice* (pp. 61-76). Dordrecht: Springer Netherlands.

Vosniadou, S. (2019). The Development of Students' Understanding of Science. *Frontiers in Education, 4*(32). https://doi.org/10.3389/feduc.2019.00032

Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 3-34). New York, NY: Routledge.

Webb, M. E., Little, D. R., & Cropper, S. J. (2018). Once more with feeling: Normative data for the aha experience in insight and noninsight problems. *Behav Res Methods, 50*(5), 2035-2056. https://doi.org/10.3758/s13428-017-0972-9

Zhong, C.-B., Dijksterhuis, A., & Galinsky, A. D. (2008). The Merits of Unconscious Thought in Creativity. *Psychological Science, 19*(9), 912-918.

Supplementary Material to "Now I get it !" : Eureka experiences during the acquisition of mathematical concepts

Charlotte Barot, Louise Chevalier, Lucie Martin, and Véronique Izard

This document has three parts. First, we present our general testing plan: how participants were assigned to the different experimental conditions, and which preliminary analyses were run as testing progressed. Second, we provide a full presentation of our paradigm, describing all the tasks that were administered to our participants. We also explain the reasons that lead us to discard some of these tasks from the analyses presented in the associated paper. Third, we present detailed results for the analyses presented in the associated paper; as well as for the same analyses applied to all the participants (including two groups not treated with a teaching phase, which were excluded from the main analyses; see 'general testing plan').

I. General testing plan

Testing was conducted in three batches. First, we administered the 1- and 7-lesson conditions in alternation, until we reached a sample size of 14 included participants in each group. The data of these two conditions was analyzed to check whether the teaching condition had an effect on test phase accuracy. We also inspected the data from the different tasks to check for any inconsistency, which led us to discard one task from analyses (see 'detailed presentation of the tasks'). Crucially, at this first stage we did not perform analyses on participants' Eureka reports in relation to performance or confidence judgments – the main analyses reported in our paper.

As a second step, we tested 28 participants (15 women, age 18-48 years, Mean = 25.6 years, exact age missing for two participants, 2-8 years of education in mathematics after 10th grade, Mean = 4.2 years) in two baseline conditions without a teaching phase. The participants of the first baseline group were presented with the one-page document with examples of great circles that served as an introduction to the lessons in the teaching groups, while the participants of the second baseline group proceeded directly from the inclusion phase to the test phase. Although we initially planned to include these two groups in the

analyses, we eventually decided to exclude them, because we realized that their inclusion blurred the conclusions that could be derived from our experiment. First, participants of the baseline groups potentially evaluated their confidence based on different criteria: for example, they could feel very confident about the idea that small circles drawn on spheres correspond to straight lines, since they were never told that small circles are not straight. Similarly, while participants in the baseline groups reported Eureka experiences, the insights they gained in these episodes could have qualitatively different contents compared to the insights of the other groups. Lastly, since the baseline groups did not receive any lessons to study, it is unclear whether they engaged in learning at all, and thus their responses may be unrelated to the question at stake in our experiment - the role of Eureka experiences in concept learning.

In the last part of this document, we present the results of our analyses with the baseline groups included. These analyses yield the same results as the analyses presented in the associated paper.

Finally, in a third step, we tested participants in the 3- and 5-lesson conditions, with participants assigned to these two conditions in alternation.

This testing plan was adopted because we suspected that it may be difficult to recruit enough participants to complete all the groups – a suspicion that proved right, as we had to stop data collection after the 3- and 5-lesson testing batch, and could not test participants with 2, 4 or 6 lessons. Yet, we wanted to be sure to be able to run meaningful analyses on our data even if we had to stop data collection at some point. Thus, across the different batches the average number of lessons studied was kept constant (an average of 4 lessons for the 1- and 7-lesson batch, the 3- and 5-lesson batch, and the 2- and 6- or 4-lesson batches that could have followed), with the exception of the baseline batch. Consequently, and given that we analyzed the effect of our teaching intervention (the only aspect of the experiment to vary across batches) by means of a numerical variable for number of lessons, differences in performance between batches could not be captured by this variable, and thus would not be spuriously attributed to the manipulation of the teaching phase.

Because the average number of lessons is different in the baseline batch, the supplementary analyses presented in the third part of this document may present biases, unlike the main analyses reported in the

associated paper. For this reason, as well as for the reasons explained above, the reader is advised to consider these results with caution.

	Baseline groups (N=28)	1-lesson group (N=14)	3-lesson group (N=14)	5-lesson group (N=14)	7-lesson group (N=14)
Planar geometry test (accuracy)	98% (85% to 100%)	95% (85% to 100%)	98% (90% to 100%)	95% (85% to 100%)	98% (85% to 100%)
Straight lines identification on the sphere, inclusion phase (accuracy)	66% (50% to 83%)	68%% (58% to 83%)	70%% (50% to 83%)	65%% (42% to 75%)	68% (50% to 83%)
Age (years)	25.6 years (18y. to 48y.)	25.7 years (18y. to 41y.)	23.7 years (20y. to 30y.)	28.2 years (19y. to 43y.)	23.9 years (20y. to 27y.)
Gender	15 Female / 13 Male	10 Female / 4 Male	11 Female / 3 Male	11 Female / 3 Male	8 Female / 6 Male
Education in mathematics (number of years after 10th grade)	4.2 years (2y. to 8y.)	4.1 years (2y. to 7y.)	3.9 years (2y. to 6y.)	3.1 years (2y. to 7y.)	4.6 years (0y. to 7y.)

Table S0. Demographic data (age, gender, education in mathematics) and average performance on inclusion tests by teaching condition (baseline, 1-, 3-, 5- or 7-lesson), for participants included in the final groups. Since the two baseline groups were analyzed together in the supplementary analyses of part III, they were pooled together in this table.

II.   Full description of the paradigm

Our paradigm comprised four phases: an *inclusion* phase, a *teaching* phase, a *test* phase, and a final *Eureka* phase where participants were asked about Eureka experiences. Here we present the comprehensive list of the tasks that were administered to the participants in each of these phases, including some tasks that were not described and analyzed in the associated paper.

Inclusion phase	Teaching phase	Test phase	Eureka phase
Definition of straight lines (1) Planar geometry test Straight lines identification on spheres	Introduction: great circles Lessons about straight lines on the sphere	Confidence judgment (1) *Subjective efficiency of the teaching phase* Straight lines identification on spheres *Definition of straight lines (2)* Straight lines identification on various surfaces Confidence judgment (2) Reasoning about straight lines on spheres and on other surfaces Confidence judgement (3) *Definition of straight lines (3)* *Retrospective ratings of understanding*	Eureka report

Table S1. Comprehensive list of all the tasks administered to the participants in the four phases of the experiment. The tasks highlighted in italics are not described in the associated paper.

Inclusion phase

Definition of straight lines. Participants were first asked to write a definition for the notion of straight line. Analyses of the definitions produced over the course of the experiment (three definitions for each participant) are not included in the associated paper. They will be reported in another paper analyzing people's intuitive concept of straight lines.

Planar geometry test. This test was adapted from Izard et al. (2011; planar geometry condition, questions 1-20)[1]. Participants were first introduced to a planar surface, extending indefinitely, on which points and straight lines could be drawn. Straight lines were described as lines that never turn, neither on the left nor on the right, and that continue straight ahead indefinitely. After this introduction, participants were asked a series of twenty illustrated questions about the properties of straight lines on this infinite plane. Questions were presented both in writing and orally through an audio recording, and participants ticked their answer (yes or no) on a response sheet. Participants were included if they made no more than 3 errors.

Straight lines identification on the sphere. In each trial, participants were presented with a photograph of a sphere (a table tennis ball) with a line drawn on it, and were asked to indicate whether the line was "straight" or not. Three types of trials were presented: non-circles (non-straight; e.g. wavy line, line looping and crossing itself to form an 8 figure), great circles (straight), and small circles (non-straight, but typically judged to be straight by most adults). Each category counted 4 trials, for a total of 12 trials presented in a randomized order. Participants responded by pressing the 'Y' key for yes ('oui') or the 'N' key for no ('non'). They were included if they made at least two mistakes on the small circle trials (i.e. they incorrectly judged small circles to be straight lines) or if they made at least two mistakes on the great circle trials (i.e. they incorrectly judged that great circles are not straight).

Teaching phase

Introduction to great circles. Participants were given a one-page document defining the "great circles" of a sphere as circles that have the same radius as the sphere on which they are drawn. This introduction also provided illustrations of great circles drawn at various orientations. A translated version of this document is accessible on the Github repository of the project (https://github.com/charlusb/Analyses_Eurekamaths).

Lessons. Participants were given 1 to 7 lessons about straight lines in spherical geometry. The lessons used simple physical models to explain why great circles correspond to straight lines on the sphere, but smaller circles do not. These models were: rolling a toy car on a ball (the car can be rolled along a great circle, but

---

[1] Izard, V., Pica, P., Spelke, E. S., and Dehaene, S. (2011). Flexible intuitions of Euclidean geometry in an Amazonian indigene group. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(24), 9782–9787. https://doi.org/10.1073/pnas.1016686108

not along a small circle; 2 lessons), applying scotch tape on a ball (the tape wrinkles if applied along a small circle, but remains smooth if applied along a great circle; 2 lessons), pinning a rubber band on a ball (the rubber band naturally aligns with a great circle; 2 lessons), and flight routes (flight routes seem curved when mapped on a planisphere, but prove to be the shortest route and a portion of a great circle when mapped on a globe; 1 lesson). A translated version of the lessons can be found on the Github repository of the project: https://github.com/charlusb/Analyses_Eurekamaths). For each teaching condition, different orders of presentation were created to ensure that each lesson appeared in each position, and that a given lesson was not always followed or preceded by the same lesson. To encourage the participants to study the lessons in depth, we asked them to write a summary of each lesson just after reading it. We also asked them to rate whether they found the lesson convincing, on a scale graded from 0 to 10. These judgments of convincingness were collected twice: a first time after reading each lesson, and a second time for all lessons together (except for the group who received only one lesson to read, as this would have resulted in asking the exact same question twice in a row).

This and other tasks of the teaching phase (see Subjective efficiency of the teaching phase below) were included as an effort to develop a measure of the temporal dynamics of the learning process. Analyses of these tasks will be reported in a different paper, together with other experiments from that line of research.

Test phase

Confidence judgment. Participants were asked to rate how much they felt they understood the notion of straight line, on a scale graduated from 0 to 10 (first rating of confidence).

Subjective efficiency of the teaching phase. Participants from the 3-, 5- and 7-lesson groups were given two questions to answer. First, they were asked whether they felt that the elements presented had helped them improve their understanding of straight lines (yes or no). Second, participants ranked the different lessons they had studied by mapping them on an oriented line, from the least convincing to the most convincing. This last task was not given to the 1-lesson group because they would have had only 1 lesson to rank. The first question about the subjective impact of the lessons on participants' understanding was missing for this group by mistake.

6

At this point, the first experimenter left and was replaced by a second experimenter who was blind to the teaching condition assigned to the participants.

Straight lines identification on spheres. This task was identical to the spherical geometry inclusion task.

Definition of straight lines. Participants were asked to write a definition for the notion of straight line (second definition).

Straight lines identification on various surfaces. Participants were presented with lines drawn on four different surfaces: cone (8 trials), cylinder (6 trials), cube (8 trials) and torus (4 trials). Each trial displayed two photographs showing a front and a back view of a surface on which a line had been drawn. Participants were asked to judge whether the line presented was straight or not and indicated their answer by keypress. The task crossed the two variables of straightness (straight, not straight) and planarity (planar, non-planar): 4 trials presented non-planar non-straight lines (corresponding to non-circle lines on the sphere), 3 trials presented planar straight lines (corresponding to great circles on the sphere), 10 trials presented planar non-straight lines (corresponding to small circles on the sphere), and 9 trials presented non-planar straight lines (there are no corresponding examples on the sphere, but such lines can exist on other surfaces). Trials were presented in a random order.

Confidence judgment. Participants were asked to indicate how much they felt they understood the notion of straight line on a 0-10 graduated scale (second rating of confidence).

Reasoning about straight lines on the sphere and on other surfaces. This task consisted in a list of mathematical assertions, which participants judged to be true or false: eight assertions about the straight lines of the sphere, followed by eight assertions about straight lines on arbitrary surfaces. The assertions were presented in a fixed order, on paper. Participants were given written definitions for the terms 'parallel' and 'perpendicular', which appeared in some of the assertions. They answered by ticking one of four response options for each assertion: 'yes - certain', 'yes - uncertain', 'no - uncertain', and 'no - certain'.

Confidence judgment. Participants indicated how much they felt they understood the notion of straight line on a 0-10 graduated scale (third rating of confidence). One participant from the 1-lesson group inadvertently failed to answer this third confidence question.

Definition of straight lines. Participants were asked to write a definition for the notion of straight line (third definition).

Retrospective judgement of confidence. Participants were asked to evaluate retrospectively their understanding of straight lines at three time points: before the teaching phase, at the end of the teaching phase and at the end of the test phase.

This task was originally included as another measure of participants' reflective introspection about their own learning: we intended to measure how much people thought that they had progressed in their understanding of straight lines during the course of the experiment. However, inspection of the data of the first two groups (1 and 7 lessons, see 'general testing plan') revealed that participants often misunderstood this task. Indeed, if they had responded as intended, i.e. by indicating how much they thought they understood the notion of generalized straight line at several stages of the experiment *in the light of the understanding they had gained at the end of the experiment*, their ratings should progressively increase, or perhaps remain stable – but they cannot decrease. Quite the contrary, we found that the ratings produced by 19 out of 28 participants expressed a decrease in understanding at some point during the experiment. This suggests that many participants misinterpreted our question as referring to their *feeling of understanding as they experienced it at different stages of the experiment* – perhaps they thought that we wanted to study whether they could faithfully remember these feelings retrospectively. We thus chose to discard this measure from our analyses, yet the task was included for all the participants, for the sake of consistency between groups.

Eureka phase

The final phase was administered by the experimenter who had been in charge of administering the teaching phase.

Eureka report. This final phase aimed at measuring whether people had experienced Eureka's during the course of the experimental session. Participants were first given a description of the sensations associated with Eureka experiences (adapted from Jung-Beeman et al, 2004[2]): episodes where a new understanding arises suddenly and unexpectedly, and is associated with a feeling of certainty. Participants indicated whether they experienced such episodes at some point during the experiment (yes or no) – this answer was used as our measure of Eureka report. Next, participants were presented with vignettes illustrating the different phases of the experiment (one vignette for each lesson, pre- and post-teaching task, confidence question, and definition of straight lines), so that they could indicate when exactly in the experiment they had experienced Eureka's. A few participants from the 3- and 5-lesson groups were also asked to describe the insights that occurred to them in these occasions.

While testing the first two groups (1 and 7 lessons), we realized that the timing reported for Eureka episodes was sometimes quite unclear. For instance, some participants could select a particular lesson because they remembered having a sudden insight *about* this lesson, while they did not remember when exactly in the course of the study they had this insight. We thus decided against including the Eureka report by time point in our main analyses, but we nonetheless chose to include this task for subsequent testing batches, as it had proven very useful to obtain informal information about participants' first-person experience in the study.

III.   Supplementary analyses

1. Detailed results of the main analyses

    a. Effect of the number of lessons on test phase performance

---

[2] Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., Reber, P. J., & Kounios, J. (2004). Neural Activity When People Solve Verbal Problems with Insight. *PLoS Biology, 2*(4), e97. https://doi.org/10.1371/journal.pbio.0020097

The logistic mixed model reported in the associated paper identified a significant interaction between number of lessons and test condition on participants' test phase accuracy ($\chi^2 = 36.2$, $p < .001$). Table S2 explores this significant interaction by examining the linear effect of number of lessons in each test condition.

Linear effect of number of lessons on accuracy in each test condition

Test Condition	Estimated Trend	Standard Error	95% Confidence Interval, Lower	95% Confidence Interval, Upper	Z-ratio	Holm corrected p-value
Non-circle lines on spheres	0.11	0.13	-0.25	0.48	0.87	1.00
Great circles on spheres	0.15	0.19	-0.37	0.67	0.79	1.00
Small circles on spheres	**0.51**	**0.09**	**0.27**	**0.75**	**5.93**	**< .001**
Non-straight non-planar lines on various surfaces (eq. non-circle lines on spheres)	0.29	0.15	-0.12	0.70	1.99	.28
Straight planar lines on various surfaces (eq. great circles on spheres)	0.03	0.12	-0.31	0.37	0.24	1.00
Non-straight planar lines on various surfaces (eq. small circles on spheres)	**0.16**	**0.05**	**0.01**	**0.31**	**3.03**	**.017**
Straight non-planar lines on various surfaces (no eq. on spheres)	0.03	0.06	-0.13	0.18	0.45	1.00
Reasoning on the sphere	**0.18**	**0.06**	**0.02**	**0.35**	**3.14**	**.014**
Reasoning on various surfaces	0.11	0.06	-0.05	0.28	1.91	.28

Table S2. Estimated linear trend for the effect of number of lessons on accuracy in each test condition. Significant trends are highlighted in bold. Results are given in logit scale.

b. Relation between Eureka experiences and test phase performance

In the associated paper, we reported a significant interaction between a variable indicating whether participants had reported a Eureka experience and the variable for test condition on participants' test phase accuracy (simple model without covariates $\chi^2 = 27.1$, $p < .001$, model with covariates for number of lessons and education in mathematics, $\chi^2 = 34.4$, $p < .001$). Table S3 explores these significant interactions by examining the estimated contrast in accuracy between participants who did vs. did not report Eureka experiences in each test condition.

Effect of Eureka report on accuracy in each test condition

Test Condition	Estimated Contrast	Standard Error	95% Confidence Interval, Lower	95% Confidence Interval, Upper	Z-ratio	Holm corrected *p*-value
Non-circle lines on spheres	0.20 *0.07*	0.60 *0.64*	-1.45 *-1.69*	1.84 *1.84*	0.33 *0.12*	1.00 *1.00*
Great circles on spheres	-1.18 *-1.75*	1.12 *1.20*	-4.29 *-5.09*	1.94 *1.58*	-1.05 *-1.46*	1.00 *1.00*
Small circles on spheres	**0.96** *0.39*	**0.35** *0.38*	**0.00** *-0.67*	**1.92** *1.45*	**2.78** *1.01*	**.043** *1.00*
Non-straight non-planar lines on various surfaces (eq. non-circle lines on spheres)	-0.90 *-1.31*	0.70 *0.74*	-2.84 *-3.38*	1.03 *0.75*	-1.29 *-1.77*	1.00 *.62*
Straight planar lines on various surfaces (eq. great circles on spheres)	-0.81 *-0.95*	0.63 *0.67*	-2.56 *-2.81*	0.94 *0.90*	-1.28 *-1.43*	1.00 *1.00*
Non-straight planar lines on various surfaces (eq. small circles on spheres)	0.25 *-0.04*	0.26 *0.26*	-0.47 *-0.77*	0.97 *0.69*	0.97 *-0.15*	1.00 *1.00*
Straight non-planar lines on various surfaces (no eq. on spheres)	**1.06** *1.22*	**0.29** *0.30*	**0.27** *0.39*	**1.86** *2.05*	**3.72** *4.09*	**.002** *< .001*
Reasoning on the sphere	0.36 *0.13*	0.27 *0.28*	-0.40 *-0.65*	1.12 *0.92*	1.32 *0.47*	1.00 *1.00*
Reasoning on various surfaces	0.12 *-0.03*	0.28 *0.29*	-0.66 *-0.83*	0.91 *0.78*	0.44 *-0.10*	1.00 *1.00*

Table S3. Estimated contrast in accuracy between participants who did vs. did not report a Eureka experience in each test condition. Plain: Simple model without covariates. *Italics: Model with covariates for number of lessons and education in mathematics.* Significant contrasts are highlighted in bold. Results are given in logit scale.

c. Relation between Eureka experiences, confidence and test phase performance

The main analyses identified a significant interaction between Eureka report and test condition (simple model without covariates $\chi^2 = 26.1$, $p = .001$, model with covariates for number of lessons and education in mathematics $\chi^2 = 34.2$, $p < .001$), and between confidence and test condition (simple model $\chi^2 = 33.0$, $p < .001$, model with covariates $\chi^2 = 25.9$, $p = .001$) on participants' test phase accuracy. Tables S4 and S5 explore these interactions.

Effect of Eureka report on accuracy in each test condition (controlling for confidence)

Test Condition	Estimated Contrast	Standard Error	95% Confidence Interval, Lower	95% Confidence Interval, Upper	Z-ratio	Holm corrected *p*-value
Non-circle lines on spheres	0.08 *0.06*	0.60 *0.64*	-1.60 *-1.71*	1.75 *1.83*	0.13 *0.10*	1.00 *1.00*
Great circles on spheres	-1.13 *-2.53*	1.26 *1.68*	-4.62 *-7.19*	2.37 *2.13*	-0.89 *-1.50*	1.00 *.93*
Small circles on spheres	0.79 *0.32*	0.36 *0.38*	-0.20 *-0.75*	1.78 *1.38*	2.22 *0.83*	.21 *1.00*
Non-straight non-planar lines on various surfaces (eq. non-circle lines on spheres)	-0.97 *-1.29*	0.70 *0.75*	-2.92 *-3.37*	0.98 *0.79*	-1.38 *-1.72*	1.00 *.68*
Straight planar lines on various surfaces (eq. great circles on spheres)	-0.77 *-0.92*	0.64 *0.68*	-2.54 *-2.80*	1.00 *0.96*	-1.20 *-1.36*	1.00 *1.00*
Non-straight planar lines on various surfaces (eq. small circles on spheres)	0.25 *-0.02*	0.27 *0.27*	-0.49 *-0.76*	0.98 *0.72*	0.93 *-0.07*	1.00 *1.00*
Straight non-planar lines on various surfaces (no eq. on spheres)	**1.15** *1.26*	**0.30** *0.30*	**0.32** *0.43*	**1.97** *2.10*	**3.87** *4.18*	**.001** *< .001*
Reasoning on the sphere	0.27 *0.10*	0.28 *0.29*	-0.51 *-0.69*	1.06 *0.89*	0.97 *0.35*	1.00 *1.00*
Reasoning on various surfaces	0.13 *-0.01*	0.29 *0.28*	-0.68 *-0.82*	0.93 *0.80*	0.43 *-0.03*	1.00 *1.00*

Table S4. Estimated contrast in accuracy between participants who did vs. did not report a Eureka experience, in each test condition, controlling for confidence. Plain: Simple model without covariates. *Italics: Model with covariates for number of lessons and education in mathematics.* Significant contrasts are highlighted in bold. Results are given in logit scale.

Linear effect of confidence on accuracy in each test condition (controlling for Eureka report)

Test Condition	Estimated Trend	Standard Error	95% Confidence Interval, Lower	95% Confidence Interval, Upper	Z-ratio	Holm corrected *p*-value
Non-circle lines on spheres	0.35 *0.32*	0.22 *0.22*	-0.24 *-0.28*	0.95 *0.92*	1.64 *1.47*	.71 *1.00*
Great circles on spheres	-2.19 *-2.51*	0.89 *1.17*	-4.67 *-5.76*	0.28 *0.74*	-2.46 *-2.14*	.11 *.29*
Small circles on spheres	0.34 *0.20*	0.13 *0.14*	-0.03 *-0.18*	0.70 *0.58*	2.54 *1.44*	.099 *1.00*
Non-straight non-planar lines on various surfaces (eq. non-circle lines on spheres)	0.18 *-0.01*	0.21 *0.22*	-0.42 *-0.62*	0.77 *0.60*	0.82 *-0.05*	1.00 *1.00*
Straight planar lines on various surfaces (eq. great circles on spheres)	-0.15 *-0.20*	0.21 *0.22*	-0.74 *-0.80*	0.44 *0.40*	-0.71 *-0.93*	1.00 *1.00*
Non-straight planar lines on various surfaces (eq. small circles on spheres)	0.00 *-0.06*	0.10 *0.09*	-0.27 *-0.32*	0.28 *0.21*	0.04 *-0.60*	1.00 *1.00*
Straight non-planar lines on various surfaces (no eq. on spheres)	-0.11 *-0.10*	0.11 *0.10*	-0.40 *-0.39*	0.18 *0.19*	-1.06 *-0.92*	1.00 *1.00*
Reasoning on the sphere	0.16 *0.09*	0.10 *0.10*	-0.13 *-0.19*	0.44 *0.37*	1.51 *0.88*	.79 *1.00*
Reasoning on various surfaces	-0.01 *-0.06*	0.11 *0.10*	-0.30 *-0.35*	0.29 *0.23*	-0.07 *-0.59*	1.00 *1.00*

Table S5. Estimated linear trend for the effect of confidence on accuracy in each test condition, controlling for Eureka report. Plain: Simple model without covariates. *Italics: Model with covariates for number of lessons and education in mathematics.* Significant trends are highlighted in bold. Results are given in logit scale.

2. Analyses including the participants who were not treated with a teaching phase

    a. Effect of the number of lessons on test phase performance

Logistic mixed model analysis of the effect of number of lessons on test phase accuracy

	df	χ^2	p
Test condition	**8**	**1003.2**	**< .001**
Number of lessons	**1**	**23.7**	**< .001**
Education in Mathematics	1	2.04	.15
Test condition*Number of lessons	**8**	**51.9**	**< .001**
Test condition*Education in mathematics	**8**	**19.0**	**.015**

Table S6. LogLik = -2395.5, Random effect (participant): variance = 0.27. Significant effects are highlighted in bold.

Linear effect of number of lessons on accuracy in each test condition

Test Condition	Estimated Trend	Standard Error	95% Confidence Interval, Lower	95% Confidence Interval, Upper	Z-ratio	Holm corrected *p*-value
Non-circle lines on spheres	0.08	0.09	-0.18	0.34	0.88	1.00
Great circles on spheres	**0.40**	**0.14**	**0.01**	**0.78**	**2.82**	**.031**
Small circles on spheres	**0.40**	**0.06**	**0.24**	**0.56**	**6.97**	**< .001**
Non-straight non-planar lines on various surfaces (eq. non-circle lines on spheres)	0.01	0.11	-0.29	0.31	0.08	1.00
Straight planar lines on various surfaces (eq. great circles on spheres)	**0.23**	**0.08**	**0.01**	**0.45**	**2.85**	**.031**
Non-straight planar lines on various surfaces (eq. small circles on spheres)	**0.09**	**0.04**	**-0.00**	**0.19**	**2.66**	**.039**
Straight non-planar lines on various surfaces (no eq. on spheres)	0.03	0.04	-0.07	0.13	0.77	1.00
Reasoning on the sphere	**0.12**	**0.04**	**0.01**	**0.22**	**3.03**	**.020**
Reasoning on various surfaces	0.09	0.04	-0.02	0.20	2.22	.10

Table S7. Estimated linear trend for the effect of number of lessons on accuracy in each test condition.

Significant trends are highlighted in bold. Results are given in logit scale.

b. Effect of the number of lessons on Eureka experiences

Logistic regression analysis of the effect of number of lessons on Eureka report

	χ^2	df	p
Number of lessons	**7.7**	**1**	**.005**
Education in mathematics	0.2	1	.63

Table S8. Significant effects are highlighted in bold.

c. Relation between Eureka experiences and test phase performance

Logistic mixed model analysis of the relation between Eureka report and test phase accuracy

	No covariate model			Covariates model		
	df	χ^2	p	df	χ^2	p
Test condition	**8**	**1017.5**	**< .001**	**8**	**1010.6**	**< .001**
Eureka report	1	4.0	.46	1	0.9	.35
Number of lessons				**1**	**20.0**	**< .001**
Education in mathematics				1	2.16	.14
Test condition*Eureka report	**8**	**16.5**	**.036**	**8**	**16.1**	**.041**
Test condition*Number of lessons				**8**	**51.9**	**< .001**
Test condition*Education in mathematics				**8**	**18.8**	**.017**

Table S9. Results of the two mixed models analyzing the relation between Eureka report and test phase accuracy. Left: simple model without covariates, right: model accounting for years of education in mathematics and number of lessons. Simple model: Loglik = -2430.6, random effect (participant): variance = 0.33. Model with covariates: LogLik = -2386.6, Random effect (participant): variance = 0.26. Significant effects are highlighted in bold.

Effect of Eureka report on accuracy in each test condition

Test Condition	Estimated contrast	Standard Error	95% Confidence Interval, Lower	95% Confidence Interval, Upper	Z-ratio	Holm corrected *p*-value
Non-circle lines on spheres	0.32 *0.22*	0.46 *0.48*	-0.96 *-1.10*	1.61 *1.55*	0.70 *0.461*	1.00 *1.00*
Great circles on spheres	**1.57** *1.18*	**0.54** *0.55*	**0.07** *-0.35*	**3.08** *2.72*	**2.89** *2.14*	**.032** *.26*
Small circles on spheres	0.56 *0.10*	0.26 *0.29*	-0.17 *-0.69*	1.29 *0.89*	2.14 *0.35*	.22 *1.00*
Non-straight non-planar lines on various surfaces (eq. non-circle lines on spheres)	-0.77 *-0.76*	0.61 *0.63*	-2.46 *-2.51*	0.93 *0.99*	-1.25 *-1.21*	1.00 *1.00*
Straight planar lines on various surfaces (eq. great circles on spheres)	0.22 *-0.14*	0.37 *0.38*	-0.80 *-1.20*	1.24 *0.93*	0.60 *-0.36*	1.00 *1.00*
Non-straight planar lines on various surfaces (eq. small circles on spheres)	0.16 *0.02*	0.19 *0.19*	-0.37 *-0.51*	0.69 *0.55*	0.83 *0.10*	1.00 *1.00*
Straight non-planar lines on various surfaces (no eq. on spheres)	**0.61** *0.64*	**0.21** *0.21*	**0.03** *0.06*	**1.19** *1.23*	**2.92** *3.05*	**.032** *.021*
Reasoning on the sphere	0.26 *0.13*	0.21 *0.21*	-0.31 *-0.44*	0.83 *0.70*	1.27 *0.63*	1.00 *1.00*
Reasoning on various surfaces	0.18 *0.06*	0.21 *0.21*	-0.41 *-0.53*	0.76 *0.64*	0.83 *0.26*	1.00 *1.00*

Table S10. Estimated contrast in accuracy between participants who did vs. did not report a Eureka experience, in each test condition. Plain: Simple model without covariates. *Italics: Model with covariates for number of lessons and education in mathematics.* Significant contrasts are highlighted in bold. Results are given in logit scale.

d. Relation between Eureka experiences and confidence

i. Correlation tests

Correlation table for confidence ratings and Eureka reports

	Confidence 1	Confidence 2	Confidence 3	Eureka
Confidence 1	X	**$\rho(84) = .67$** **$p < .001$**	**$\rho(83) = .64$** **$p < .001$**	$\rho(84) = .20$ $p = .17$
Confidence 2	**$\rho(82) = .64$** **$p < .001$**	X	**$\rho(83) = .86$** **$p < .001$**	$\rho(84) = .19$ $p = .17$
Confidence 3	**$\rho(81) = .65$** **$p < .001$**	**$\rho(81) = .87$** **$p < .001$**	X	$\rho(83) = .21$ $p = .17$
Eureka	$\rho(82) = .12$ $p = .39$	$\rho(82) = .15$ $p = .39$	$\rho(81) = .21$ $p = .19$	X

Table S11. Spearman's ρ coefficients and *p*-values for pairwise correlation tests. Above diagonal: without covariates, below: with number of lessons and years of education in mathematics as covariates. Significant correlations are highlighted in bold. All *p*-values were corrected for multiple comparisons using Holm's method (applied separately for the analyses with and without covariates). Note that the third rating of confidence was missing for one participant in the 1-lesson group, hence the difference in degrees of freedom. Confidence 1: measured just after participants completed the teaching phase; Confidence 2: measured after the various surfaces straight lines identification task; Confidence 3: measured after the reasoning task.

ii. Relation between Eureka experiences, confidence and test phase performance

Logistic mixed model analysis of the relation between Eureka report, confidence and test phase accuracy

	No covariate model			Covariates model		
	df	χ^2	p	df	χ^2	p
Test condition	**8**	**1025.6**	**< .001**	**8**	**1020.0**	**< .001**
Eureka report	**1**	**4.4**	**.036**	1	1.6	.21
Confidence	1	0.6	.44	1	3.1	.078
Number of lessons				**1**	**21.3**	**< .001**
Education in mathematics				1	2.2	.14
Test condition*Eureka report	**8**	**19.2**	**.014**	**8**	**18.6**	**.017**
Test condition*Confidence	**8**	**17.6**	**.024**	8	14.1	.080
Test condition*Number of lessons				**8**	**48.1**	**< .001**
Test condition*Education in mathematics				**8**	**18.9**	**.015**

Table S12. Results of the two mixed models analyzing the relations of Eureka report and confidence to test phase accuracy. Left: simple model without covariates, right: model with covariates for years of education in mathematics and number of lessons. Simple model: LogLik = -2421.7, Random effect (participant): variance = 0.33, Covariates model: LogLik = -2378.6, Random effect (participant): variance = 0.25. Significant effects are highlighted in bold.

Effect of Eureka report on accuracy in each test condition (controlling for confidence)

Test Condition	Estimated Contrast	Standard Error	95% Confidence Interval, Lower	95% Confidence Interval, Upper	Z-ratio	Holm corrected *p*-value
Non-circle lines on spheres	0.26 *0.20*	0.47 *0.48*	-1.03 *-1.13*	1.55 *1.53*	0.56 *0.42*	1.00 *1.00*
Great circles on spheres	**1.83** *1.51*	**0.56** *0.59*	**0.27** *-0.11*	**3.39** *3.13*	**3.25** *2.58*	**.010** *.080*
Small circles on spheres	0.48 *0.09*	0.27 *0.29*	-0.26 *-0.70*	1.22 *0.89*	1.79 *0.70*	.51 *.89*
Non-straight non-planar lines on various surfaces (eq. non-circle lines on spheres)	-0.77 *-0.74*	0.62 *0.63*	-2.47 *-2.49*	0.94 *1.01*	-1.25 *-1.18*	1.00 *1.00*
Straight planar lines on various surfaces (eq. great circles on spheres)	0.26 *-0.08*	0.37 *0.39*	-0.78 *-1.16*	1.29 *0.99*	0.69 *-0.21*	1.00 *1.00*
Non-straight planar lines on various surfaces (eq. small circles on spheres)	0.15 *0.03*	0.20 *0.19*	-0.39 *-0.51*	0.70 *0.56*	0.79 *0.13*	1.00 *1.00*
Straight non-planar lines on various surfaces (no eq. on spheres)	**0.67** *0.68*	**0.21** *0.21*	**0.08** *0.09*	**1.27** *1.27*	**3.13** *3.21*	**.014** *.012*
Reasoning on the sphere	0.27 *0.16*	0.21 *0.21*	-0.31 *-0.42*	0.86 *0.73*	1.31 *0.75*	1.00 *1.00*
Reasoning on various surfaces	0.20 *0.08*	0.21 *0.21*	-0.40 *-0.50*	0.79 *0.67*	0.92 *0.40*	1.00 *1.00*

Table S13. Estimated contrast in accuracy between participants who did vs. did not report a Eureka experience, in each test condition, controlling for confidence. Plain: Simple model without covariates. *Italics: Model with covariates for number of lessons and education in mathematics.* Significant contrasts are highlighted in bold. Results are given in logit scale.

Linear effect of confidence on accuracy in each test condition (controlling for Eureka report)

Test Condition	Estimated Trend	Standard Error	95% Confidence Interval, Lower	95% Confidence Interval, Upper	Z-ratio	Holm corrected *p*-value
Non-circle lines on spheres	0.16 *0.14*	0.13 *0.13*	-0.20 *-0.23*	0.52 *0.51*	1.25 *1.05*	1.00 *1.00*
Great circles on spheres	-0.51 ***-0.57***	0.19 ***0.18***	-1.04 ***-1.08***	0.03 ***-0.06***	-2.64 ***-3.07***	.074 ***.019***
Small circles on spheres	0.13 *0.01*	0.08 *0.08*	-0.08 *-0.22*	0.35 *0.23*	1.70 *0.09*	.71 *1.00*
Non-straight non-planar lines on various surfaces (eq. non-circle lines on spheres)	0.04 *-0.00*	0.16 *0.17*	-0.42 *-0.48*	0.49 *0.48*	0.24 *-0.00*	1.00 *1.00*
Straight planar lines on various surfaces (eq. great circles on spheres)	-0.07 *-0.12*	0.12 *0.12*	-0.39 *-0.44*	0.26 *0.20*	-0.57 *-1.05*	1.00 *1.00*
Non-straight planar lines on various surfaces (eq. small circles on spheres)	0.01 *-0.02*	0.06 *0.06*	-0.15 *-0.17*	0.17 *0.13*	0.23 *-0.36*	1.00 *1.00*
Straight non-planar lines on various surfaces (no eq. on spheres)	-0.08 *-0.09*	0.06 *0.06*	-0.25 *-0.25*	0.09 *0.08*	-1.30 *-1.42*	1.00 *1.00*
Reasoning on the sphere	-0.02 *-0.07*	0.06 *0.06*	-0.19 *-0.23*	0.15 *0.10*	-0.26 *-1.05*	1.00 *1.00*
Reasoning on various surfaces	-0.04 *-0.07*	0.06 *0.06*	-0.21 *-0.24*	0.14 *0.10*	-0.57 *-1.14*	1.00 *1.00*

Table S14. Estimated linear trend for the effect of confidence on accuracy, in each test condition, controlling for Eureka report. Plain: Simple model without covariates. *Italics: Model with covariates for number of lessons and education in mathematics.* Significant trends are highlighted in bold. Results are given in logit scale.