

1 **Geometric models reveal the hidden structure of**
2 **conceptual knowledge**

3 Paxton C. Fitzpatrick and Jeremy R. Manning^{*}

Dartmouth College

*Corresponding author: jeremy.r.manning@dartmouth.edu

4 **Abstract**

5 We develop a mathematical framework, based on natural language processing models, for
6 tracking and characterizing the acquisition of conceptual knowledge. Our approach embeds each
7 concept in a high dimensional representation space, where nearby coordinates reflect similar or
8 related concepts. We tested our approach using behavioral data collected from a group of
9 college students. In the experiment, we asked the participants to answer sets of quiz questions
10 interleaved between watching two course videos from the Khan Academy platform. We applied
11 our framework to the videos' transcripts, and to text of the quiz questions, to quantify the
12 content of each moment of video and each quiz question. We used these embeddings, along with
13 participants' quiz responses, to track how the learners' knowledge changed after watching each
14 video. Our findings show how a limited set of quiz questions may be used to construct rich and
15 meaningful representations of what each learner knows, and how their knowledge changes over
16 time as they learn.

17 **Keywords:** education, learning, knowledge, concepts, natural language processing

¹⁸ **Introduction**

¹⁹ How do we acquire conceptual knowledge? Memorizing course lectures or textbook chapters by
²⁰ rote can lead to the superficial *appearance* of understanding the underlying content, but achieving
²¹ true conceptual understanding seems to require something deeper and richer. Does conceptual
²² understanding entail connecting newly acquired information to the scaffolding of one's existing
²³ knowledge or experience [1, 5, 7, 8, 22]? Or weaving a lecture's atomic elements (e.g., its compo-
²⁴ nent words) into a structured network that describes how those individual elements are related?
²⁵ Conceptual understanding could also involve building a mental model that transcends the mean-
²⁶ ings of those individual atomic elements by reflecting the deeper meaning underlying the gestalt
²⁷ whole [14, 16, 21].

²⁸ The difference between “understanding” and “memorizing,” as framed by the researchers
²⁹ in education, cognitive psychology, and cognitive neuroscience [9, 10, 13, 16, 21] has profound
³⁰ analogs in the fields of natural language processing and natural language understanding. For
³¹ example, considering the raw contents of a document (e.g., its constituent symbols, letters, and
³² words) might provide some information about what the document is about, just as memorizing
³³ a passage might be used to answer simple questions about the passage [e.g., whether it might
³⁴ contain words related to furniture versus physics; 2, 3, 15]. However, modern natural language
³⁵ processing models [e.g., 4, 6, 20] also attempt to capture the deeper meaning *underlying* those
³⁶ atomic elements. These models consider not only the co-occurrences of those elements within
³⁷ and across documents, but also patterns in how those elements appear across different scales (e.g.,
³⁸ sentences, paragraphs, chapters, etc.), the temporal and grammatical properties of the elements,
³⁹ and other high-level characteristics of how they are used [17, 18]. According to these models, the
⁴⁰ deep conceptual meaning of a document may be captured by a feature vector in a high-dimensional
⁴¹ representation space, where nearby vectors reflect conceptually related documents. A model that
⁴² succeeds at capturing an analog of “understanding” is able to assign nearby feature vectors to
⁴³ two conceptually related documents, *even when the words contained in those documents have very little*
⁴⁴ *overlap*.

45 What form might the representation of the sum total of a person’s knowledge take? First,
46 we might require a means of systematically describing or representing the nearly infinite set of
47 possible things a person could know. Second, we might want to account for potential associations
48 between different concepts. For example, the concepts of “fish” and “water” might be associated in
49 the sense that fish live in water. Third, knowledge may have a critical dependency structure, such
50 that knowing about a particular concept might require first knowing about a set of other concepts.
51 For example, understanding the concept of a fish swimming in water first requires understanding
52 what fish and water *are*. Fourth, as we learn, our “current state of knowledge” should change
53 accordingly. Learning new concepts should both update our characterizations of “what is known”
54 and should also unlock any now-satisfied dependencies of that newly learned concept so that they
55 are “tagged” as available for future learning.

56 Here we develop a framework for modelling how knowledge is acquired during learning. The
57 central idea is to use text embedding models to define the coordinate systems of two maps: (a) a
58 *knowledge map* that describes the extent to which each concept is currently known and (b) a *learning*
59 *map* that describes the extent to which each concept could be learned. Each location on these maps
60 represents a single concept, and the geometries are defined such that related concepts are located
61 nearby in space. We use this framework to analyzing and interpreting behavioral data collected
62 from an experiment that has participants watch and answer conceptual questions about a series of
63 recorded course lectures.

64 Our primary research goal is to advance our understanding of what it means to acquire deep
65 real-world conceptual knowledge. Traditional laboratory approaches to studying learning and
66 memory (e.g., list learning studies) often draw little distinction between memorization and under-
67 standing. Instead, these studies typically focus on whether information is effectively encoded or
68 retrieved, rather than whether the information is *understood*. Approaches to studying conceptual
69 learning, such as category learning experiments, can start to investigate the distinction between
70 memorization and understanding, often by training participants to distinguish arbitrary or ran-
71 dom features in otherwise meaningless categorized stimuli. However the objective of real-world
72 training, or learning from life experiences more generally, is often to develop new knowledge

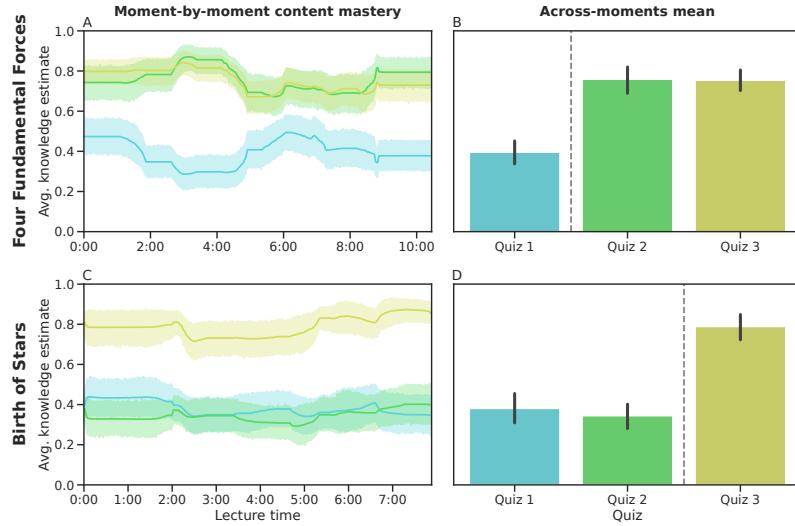


Figure 1: Estimating moment-by-moment knowledge acquisition. **A. Moment-by-moment knowledge about the *Four Fundamental Forces*.** Each trace displays the weighted proportion of correctly answered questions about the content reflected in each moment of the lecture (see *Estimating dynamic knowledge traces*), using responses from one quiz (color). The traces are averaged across participants. **B. Average estimated knowledge about the *Four Fundamental Forces*.** Each bar displays the across-timepoint average knowledge, estimated using the responses to one quiz's questions. **C. Moment-by-moment knowledge about the *Birth of Stars*.** The panel is in the same format as Panel A, but here the knowledge estimates are for the moment-by-moment content of the *Birth of Stars* lecture. **D. Average estimated knowledge about the *Birth of Stars*.** The panel is in the same format as Panel B, but here the knowledge estimates are for the content of the *Birth of Stars* lecture. All panels: error ribbons and error bars denote 95% confidence intervals, estimated across participants.

73 that may be applied in *useful* ways in the future. In this sense, the gap between modern learning
 74 theories and modern pedagogical approaches and classroom learning strategies is enormous: most
 75 of our theories about *how* people learn are inspired by experimental paradigms and models that
 76 have only peripheral relevance to the kinds of learning that students and teachers actually seek.
 77 To help bridge this gap, our study uses course materials from real online courses to inform, fit, and
 78 test models of real-world conceptual learning.

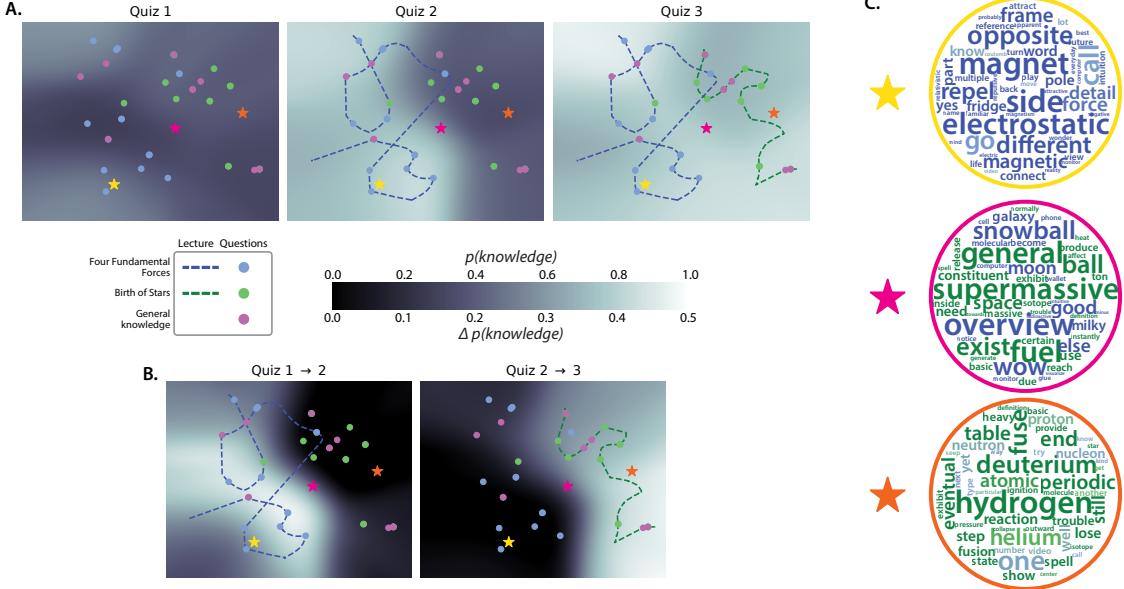


Figure 2: Mapping out the geometry of knowledge and learning. **A. Average “knowledge maps” estimated using each quiz.** Each map displays a 2D projection of the estimated knowledge about the content reflected by all regions of topic space (see *Creating knowledge and learning map visualizations*). The topic trajectories of each lecture and the coordinates of each question are indicated by dotted lines and dots. Each map reflects an average across all participants. For individual participants’ maps, see Figures S1, S2, and S3. **B. Average “learning maps” estimated between each successive pair of quizzes.** The learning maps are in the same general format as the knowledge maps in Panel A, but each coordinate in the learning maps indicates the difference between the corresponding coordinates in the indicated pair of knowledge maps—i.e., how much the estimated knowledge “changed” across the two quizzes. Each map reflects an average across all participants. For individual participants’ maps, see Figures S4 and S5. **C. Word clouds for sampled points in topic space.** Each word cloud displays the relative weights of each word reflected by the blend of topics represented at the locations of the stars in the maps. The words’ colors indicate how much each word is weighted on average across all timepoints’ topic vectors in the *Four Fundamental Forces* (blue) and *Birth of Stars* (green) videos, respectively.

⁷⁹ **Results**

⁸⁰ **Discussion**

⁸¹ **Materials and methods**

⁸² **Participants**

⁸³ We enrolled a total of 50 Dartmouth undergraduate students in our study. Participants received
⁸⁴ course credit for enrolling. We asked each participant to fill out a demographic survey that included
⁸⁵ questions about their age, gender, native spoken language, ethnicity, race, hearing, color vision,
⁸⁶ sleep, coffee consumption, level of alertness, and several aspects of their educational background
⁸⁷ and prior coursework.

⁸⁸ Participants' ages ranged from 18 to 22 years (mean: 19.52 years; standard deviation: 1.09
⁸⁹ years). A total of 15 participants reported their gender as male and 35 participants reported their
⁹⁰ gender as female. A total of 49 participants reported their native language as "English" and 1
⁹¹ reported having another native language. A total of 47 participants reported their ethnicity as
⁹² "Not Hispanic or Latino" and three reported their ethnicity as "Hispanic or Latino." Participants
⁹³ reported their races as White (32 participants), Asian (14 participants), Black or African American
⁹⁴ (5 participants), American Indian or Alaska Native (1 participant), and Native Hawaiian or Other
⁹⁵ Pacific Islander (1 participant). (Note that some participants selected multiple racial categories.)

⁹⁶ A total of 49 participants reporting having normal hearing and 1 participant reported having
⁹⁷ some hearing impairment. A total of 49 participants reported having normal color vision and 1
⁹⁸ participant reported being color blind. Participants reported having had, on the night prior to
⁹⁹ testing, 2 – 4 hours of sleep (1 participant), 4 – 6 hours of sleep (9 participants), 6 – 8 hours of sleep
¹⁰⁰ (35 participants), or 8+ hours of sleep (5 participants). They reported having consumed, on the
¹⁰¹ same day and leading up to their testing session, 0 cups of coffee (38 participants), 1 cup of coffee
¹⁰² (10 participants), 3 cups of coffee (1 participant), or 4+ cups of coffee (1 participant).

¹⁰³ No participants reported that their focus was currently impaired (e.g., by drugs or alcohol).

¹⁰⁴ Participants reported their current level of alertness, and we converted their responses to numerical
¹⁰⁵ scores as follows: “very sluggish” (-2), “a little sluggish” (-1), “neutral” (0), “fairly alert” (1), and
¹⁰⁶ “very alert” (2). Across all participants, a range of alertness levels were reported (range: -2 – 1;
¹⁰⁷ mean: -0.10; standard deviation: 0.84).

¹⁰⁸ Participants reported their undergraduate major(s) as Social Sciences (28 participants), Natural
¹⁰⁹ sciences (16), Professional (e.g., pre-med or pre-law; 8 participants), Mathematics and engineering
¹¹⁰ (7 participants), Humanities (4 participants), or Undecided (3 participants). Note that some par-
¹¹¹ ticipants selected multiple categories for their undergraduate major. We also asked participants
¹¹² about the courses they had taken. In total, 46 participants reported having taken at least one Khan
¹¹³ academy course in the past or being familiar with the Khan academy, and 4 reported not having
¹¹⁴ taken any Khan academy courses. Of the participants who reported having watched at least one
¹¹⁵ Khan academy course, 1 participant declined to report the number of courses they had watched;
¹¹⁶ 7 participants reported having watched 1–2 courses; 11 reported having watched 3–5 courses; 8
¹¹⁷ reported having watched 5–10 courses; and 19 reported having watched 10 or more courses. We
¹¹⁸ also asked participants about the specific courses they had watched, categorized under different
¹¹⁹ subject areas. In the “Mathematics” area participants reported having watched videos on AP
¹²⁰ Calculus AB (21 participants), Precalculus (17 participants), Algebra 2 (14 participants), AP Cal-
¹²¹ culus BC (12 participants), Trigonometry (11 participants), Algebra 1 (10 participants), Geometry
¹²² (8 participants), Pre-algebra (7 participants), Multivariable Calculus (5 participants), Differential
¹²³ Equations (5 participants), Statistics and Probability (4 participants), AP Statistics (2 participants),
¹²⁴ Linear Algebra (2 participants), Early Math (1 participant), Arithmetic (1 participant), and other
¹²⁵ videos not listed in our survey (6 participants). In the “Science and engineering” area participants
¹²⁶ reported having watched videos on Chemistry, AP Chemistry, or Organic Chemistry (21 partic-
¹²⁷ ipants); Physics, AP Physics I, or AP Physics II (15 participants); Biology, AP Biology; or High
¹²⁸ school Biology (15 participants); Health and Medicine (1 participant); or other videos not listed in
¹²⁹ our survey (20 participants). We also asked participants if they had specifically seen the videos
¹³⁰ used in our experiment. When we asked about the *Four Fundamental Forces* video, 45 participants
reported not having watched it before, 1 participant reported that they were not sure if they had

¹³² watched it before, and 4 participants declined to respond. When we asked about the *Birth of*
¹³³ *Stars* video, 46 participants reported not having watched it before and 4 participants declined to
¹³⁴ respond. When we asked participants about non-Khan academy online courses, they reported
¹³⁵ having watched or taken courses on Mathematics (15 participants), Science and engineering (11
¹³⁶ participants), Test preparation (9 participants), Economics and finance (3 participants), Arts and
¹³⁷ humanities (2 participants), Computing (2 participants), and other categories not listed in our
¹³⁸ survey (18 participants). Finally, we asked participants about in-person courses they had taken in
¹³⁹ different subject areas. They reported taking courses in Mathematics (39 participants), Science and
¹⁴⁰ engineering (38 participants), Arts and humanities (35 participants), Test preparation (27 partici-
¹⁴¹ pants), Economics and finance (26 participants), Computing (15 participants), College and careers
¹⁴² (7 participants), or other courses not listed in our survey (6 participants).

¹⁴³ **Experiment**

¹⁴⁴ We hand-selected two roughly 10-minute course videos from the Khan Academy platform: *The*
¹⁴⁵ *Four Fundamental Forces* (an introduction to gravity, electromagnetism, the weak nuclear force, and
¹⁴⁶ the strong nuclear force; duration: 10 minutes and 29 seconds) and *Birth of Stars* (an introduction
¹⁴⁷ to how stars are formed; duration: 7 minutes and 57 seconds). We hand-wrote 39 multiple
¹⁴⁸ choice questions: 15 about the conceptual content of *The Four Fundamental Forces*, another 15 about
¹⁴⁹ the conceptual content of *Birth of Stars*, and 9 other questions that tested for general conceptual
¹⁵⁰ knowledge about basic physics (covering material that was not presented in either video). The full
¹⁵¹ set of questions may be found in Table S1.

¹⁵² Participants began the main experiment by answering a battery of 13 randomly selected ques-
¹⁵³ tions (chosen from the full set of 39). Then they watched the *The Four Fundamental Forces* video.
¹⁵⁴ Next, they answered a second set of 13 questions (chosen at random from the remaining 26 ques-
¹⁵⁵ tions). Fourth, participants watch the *Birth of Stars* video, and finally they answered the remaining
¹⁵⁶ 13 questions. Our experimental procedure is diagramed in Figure 3. We used the experiment to
¹⁵⁷ develop and test our computational framework for estimating knowledge and learning maps.

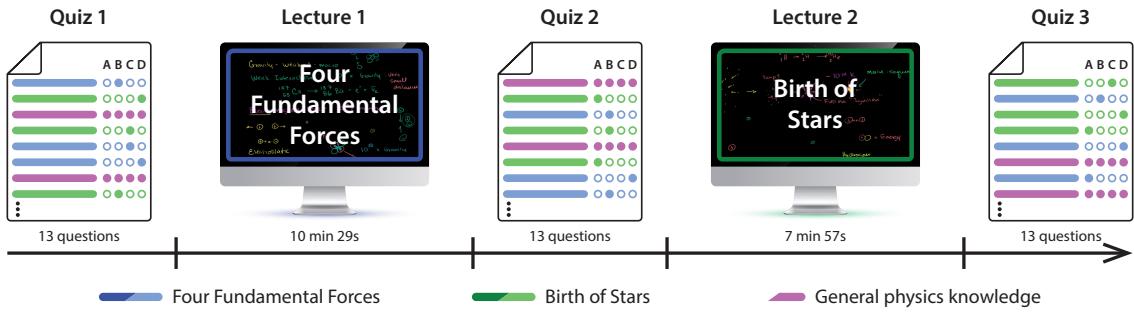


Figure 3: Experimental paradigm. Participants alternate between answering 13-question multiple choice quizzes and watching two Khan academy videos. Each quiz contains a mix of 5 questions about lecture 1, 5 questions about lecture 2, and 3 general physics knowledge questions. The specific questions reflected on each quiz, and the orders of each quiz’s questions, were randomized across participants.

158 Analysis

159 Constructing text embeddings of multiple videos and questions

160 We extended an approach developed by [12] to construct text embeddings for each moment of each
 161 lecture, and of each question in our pool. Briefly, our approach uses a topic model [3], trained on a
 162 set of documents, to discover a set of k “topics” or “themes.” Formally, each topic is defined as a set
 163 of weights over each word in the model’s vocabulary (i.e., the union of all unique words, across all
 164 documents, excluding “stop words.”). Conceptually, each topic is intended to give larger weights
 165 to set of words that appear conceptually related or that tend to co-occur in the same documents.
 166 After fitting a topic model, each document in the training set, or any *new* document that contains at
 167 least some of the words in the model’s vocabulary, may be represented as a k -dimensional vector
 168 describing how much the document (most probably) reflects each topic. (Unless, otherwise noted,
 169 we used $k = 15$ topics.)

170 As illustrated in Figure 4A, we start by building up a corpus of documents using overlapping
 171 sliding windows that span each video’s transcript. Khan Academy videos are hosted on the
 172 YouTube platform, and all YouTube videos are run through Google’s speech-to-text API [11] to
 173 derive a timestamped transcript of any detected speech in the video. The resulting transcripts
 174 contain one timestamped row per line, and each line generally corresponds to a few seconds of
 175 spoken content from the video. We defined a sliding window length of (up to) $w = 30$ transcript

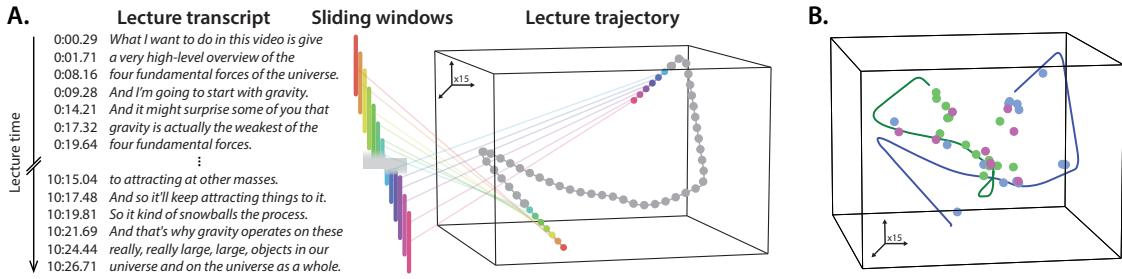


Figure 4: Constructing video content *trajectories*. **A. Building a document pool from sliding windows**. We decompose each video’s transcript into a series of overlapping sliding windows. The set of transcript snippets (across all windows) may be treated as a set of “documents” for training a text embedding model. After training a text embedding model using the two videos’ sliding windows, along with the text from each question in our pool (Tab. S1), we construct “trajectories” through text embedding space by joining the embedding coordinates of successive sliding windows from each video. **B. Embedding multiple videos and questions.** Applying the same text embedding approach to each video, along with the text of each question, results in one trajectory per video and one embedding coordinate (dot) per question (blue: Four Fundamental Forces; green: Birth of Stars; pink: general physics knowledge). Here we have projected the 15-dimensional embeddings into a 3D space using Uniform Manifold Approximation and Projection [UMAP; 19].

176 lines, and we assigned each window a timestamp according to the midpoint between its first
 177 and last lines’ timestamps. These sliding windows ramped up and down in length at the very
 178 beginning and end of the transcript, respectively. In other words, the first sliding window covered
 179 only the first line from the transcript; the second sliding window covered the first two lines; and
 180 so on. This insured that each line of the transcript appeared in the same number (w) of sliding
 181 windows. We treated the text from each sliding window as a single “document,” and we combined
 182 these documents across the two videos’ windows to create a single training corpus for the topic
 183 model. The top words from each of the 15 discovered topics may be found in Table S2.

184 After fitting a topic model to each videos’ transcripts, we could use the trained model to
 185 transform arbitrary (potentially new) documents into k -dimensional topic vectors. A convenient
 186 property of these topic vectors is that documents that reflect similar blends of topics (i.e., documents
 187 that reflect similar themes, according to the model) will yield similar (in terms of Euclidean distance,
 188 correlation, etc.) topic vectors. In general, the similarity between different documents’ topic vectors
 189 may be used to characterize the similarity in content between the documents.

190 We transformed each sliding window’s text into a topic vector, and then used linear interpo-

lation (independently for each topic dimension) to resample the resulting timeseries to once per second. This yielded a single topic vector for each second of each video. We also used the fitted model to obtain topic vectors for each question in our pool (Tab. S1). Taken together, we obtained a *trajectory* for each video, describing its path through topic space, and a single coordinate for each question (Fig. 4B). Embedding both videos and all of the questions using a common model enables us to compare the content from different moments of videos, compare the content across videos, and estimate potential associations between specific questions and specific moments of video.

198 Estimating dynamic knowledge traces

199 We used the following equation to estimate each participant's knowledge about timepoint t of a
 200 given lecture, $\hat{k}(t)$:

$$\hat{k}(t) = \frac{\sum_{i \in \text{correct}} \text{ncorr}(t, i)}{\sum_{j=1}^N \text{ncorr}(t, j)}, \quad (1)$$

201 where

$$\text{ncorr}(x, y) = \frac{\text{corr}(x, y) - \text{mincorr}}{\text{maxcorr} - \text{mincorr}}, \quad (2)$$

202 and where mincorr and maxcorr are the minimum and maximum correlations between any lecture
 203 timepoint and question, taken over all timepoints and questions across both lectures and all three
 204 question sets.

205 Intuitively, $\text{ncorr}(x, y)$ is the correlation between two topic vectors (e.g., the topic vector from one
 206 timepoint in a lecture, x , and the topic vector for one question, y), normalized by the minimum and
 207 maximum correlations (across all timepoints and questions) to range between 0 and 1, inclusive.
 208 Equation 1 then computes the weighted average proportion of correctly answered questions about
 209 the content presented at timepoint t , where the weights are given by the normalized correlations
 210 between timepoint t 's topic vector and the topic vectors for each question. The normalization
 211 step (i.e., using ncorr instead of the raw correlations) insures that every question (except the
 212 least-relevant question) contributes some non-zero amount to the knowledge estimate.

213 **Estimating held-out conceptual knowledge**

214 **Creating knowledge and learning map visualizations**

215 An important feature of our approach is that, given a trained text embedding model and partic-
216 ipants’ quiz performance on each question, we can estimate their knowledge about *any* content
217 expressable by the embedding model– not solely the content explicitly probed by the quiz ques-
218 tions. To visualize these estimates (Figs. 2, S1, S2, S3, S4, and S5), we used UMAP [19] to define a
219 2D projection of the text embedding space. Sampling the original 100-dimensional space at high
220 resolution to obtain an adequate set of topic vectors spanning the embedding space would be
221 computationally intractable. However, sampling a 2D grid is much more feasible. We defined a
222 rectangle enclosing the 2D projections of the lectures’ and quizzes’ embeddings, and we sampled
223 points from a regular 100×100 grid of coordinates that evenly tiled the enclosing rectangle. We
224 sought to estimate participants’ knowledge (and learning– i.e., changes in knowledge) at each of
225 the resulting 10000 coordinates.

226 To generate our estimates, we placed a set of 39 radial basis functions (RBFs) throughout the
227 embedding space, centered on the 2D projections for each question (i.e., we included one RBF for
228 each question). At coordinate x , the value of an RBF centered on a question’s coordinate μ , is given
229 by:

$$\text{RBF}(x, \mu, \lambda) = \exp\left\{-\frac{\|x - \mu\|^2}{\lambda}\right\}. \quad (3)$$

230 The λ term in the RBF equation controls the “smoothness” of the function, where larger values
231 of λ result in smoother maps. In our implementation we used $\lambda = 50$. Next, we estimated the
232 “knowledge” at each coordinate, x , using:

$$\hat{k}(x) = \frac{\sum_{i \in \text{correct}} \text{RBF}(x, q_i, \lambda)}{\sum_{j=1}^N \text{RBF}(x, q_j, \lambda)}. \quad (4)$$

233 Intuitively, Equation 4 computes the weighted proportion of correctly answered questions, where
234 the weights are given by how nearby (in the 2D space) each question is to the x . We also defined
235 *learning maps* as the coordinate-by-coordinate differences between any pair of knowledge maps.

236 Intuitively, learning maps reflect the *change* in knowledge across two maps.

237 References

- 238 [1] Blaye, A., Bernard-Peyron, V., Paour, J.-L., and Bonthoux, F. (2006). Category flexibility in chil-
239 dren: distinguishing response flexibility from conceptual flexibility; the protracted development
240 of taxonomic representations. *European Journal of Developmental Psychology*, 3(2):163–188.
- 241 [2] Blei, D. M. and Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the International*
242 *Conference on Machine Learning*, pages 113–120, New York, NY. Association for Computing
243 Machinery.
- 244 [3] Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine*
245 *Learning Research*, 3:993–1022.
- 246 [4] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A.,
247 Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child,
248 R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M.,
249 Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei,
250 D. (2020). Language models are few-shot learners. *arXiv*, 2005.14165.
- 251 [5] Caramazza, A. and Mahon, B. Z. (2003). The organization of conceptual knowledge: the
252 evidence from category-specific semantic deficits. *Trends in Cognitive Sciences*, 7(8):354–361.
- 253 [6] Cer, D., Yang, Y., Kong, S. Y., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-
254 Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., and Kurzweil, R. (2018). Universal
255 sentence encoder. *arXiv*, 1803.11175.
- 256 [7] Constantinescu, A. O., O'Reilly, J. X., and Behrens, T. E. J. (2016). Organizing conceptual
257 knowledge in humans with a gridlike code. *Science*, 352(6292):1464–1468.
- 258 [8] Deacon, D., Grose-Fifer, J., Yang, C. M., Stanick, V., Hewitt, S., and Dynowska, A. (2004).

- 259 Evidence for a new conceptualization of semantic representation in the left and right cerebral
260 hemispheres. *Cortex*, 40(3):467–478.
- 261 [9] Gallagher, J. J. (2000). Teaching for understanding and application of science knowledge. *School*
262 *Science and Mathematics*, 100(6):310–318.
- 263 [10] Hall, R. and Greeno, J. (2008). *21st century education: A reference handbook*, chapter Conceptual
264 learning, pages 212–221. Sage Publications.
- 265 [11] Halpern, Y., Hall, K. B., Schogol, V., Riley, M., Roark, B., Skobeltsyn, G., and Bäuml, M. (2016).
266 Contextual prediction models for speech recognition. In *Interspeech*, pages 2338–2342.
- 267 [12] Heusser, A. C., Fitzpatrick, P. C., and Manning, J. R. (2021). Geometric models reveal be-
268 havioral and neural signatures of transforming naturalistic experiences into episodic memories.
269 *Nature Human Behavior*, 5:905–919.
- 270 [13] Katona, G. (1940). *Organizing and memorizing: studies in the psychology of learning and teaching*.
271 Columbia University Press.
- 272 [14] Kintsch (1970). *Learning, memory, and conceptual processes*. Wiley.
- 273 [15] Landauer, T. K. and Dumais, S. T. (1997). A solution to Plato’s problem: the latent semantic
274 analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*,
275 104:211–240.
- 276 [16] Macellan, E. (2005). Conceptual learning: the priority for higher education. *British Journal of*
277 *Educational Studies*, 53(2):129–147.
- 278 [17] Manning, J. R. (2020). Context reinstatement. In Kahana, M. J. and Wagner, A. D., editors,
279 *Handbook of Human Memory*. Oxford University Press.
- 280 [18] Manning, J. R. (2021). Episodic memory: mental time travel or a quantum “memory wave”
281 function? *Psychological Review*, 128(4):711–725.

- 282 [19] McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform manifold approximation and
283 projection for dimension reduction. *arXiv*, 1802(03426).
- 284 [20] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word represen-
285 tations in vector space. *arXiv*, 1301.3781.
- 286 [21] Scott, P., Asoko, H., and Leach, J. (2007). *Handbook of research on science education*, chapter
287 Student conceptions and conceptual learning in science. Routledge.
- 288 [22] Simon, M. A., Tzur, R., Heinz, K., and Kinzel, M. (2004). Explicating a mechanism for
289 conceptual learning: elaborating the construct of reflective abstraction. *Journal for Research in
290 Mathematics Education*, 35(5):305–329.