Chaining and the process of scientific innovation

Emmy Liu (me.liu@mail.utoronto.ca)
Computer Science and Cognitive Science Programs
University of Toronto

Yang Xu (yangxu@cs.toronto.edu)
Department of Computer Science, Cognitive Science Program
University of Toronto

Abstract

A scientist's academic pursuit can follow a winding path. Starting with one topic of research in earlier career, one may later pursue topics that relate remotely to the initial point. Philosophers and cognitive scientists have proposed theories about how science has developed, but their emphasis is typically not on explaining the processes of innovation in individual scientists. We examine regularity in the emerging order of a scientist's publications over time. Our basic premise is that scientific papers should emerge in non-arbitrary ways that tend to follow a process of chaining, whereby novel papers are linked to existing papers with closely related ideas. We evaluate this proposal with a set of probabilistic models on the historical publications from 70 Turing Award winners. We show that an exemplar model of chaining best explains the data among the alternative models, mirroring recent findings on chaining in the growth of linguistic meaning.

Keywords: scientific innovation; chaining; exemplar model

Introduction

A scientist's academic pursuit can follow a winding path. Newton studied optics earlier in his career but later developed ideas on the laws of physics and calculus. Turing initially published on computability theory but later contributed to machine intelligence and the chemical basis of morphogenesis. How do scientists conceive ideas over time, and to what extent do earlier ideas influence the development of future ideas? Here we present a computational approach to exploring regularity in the process of scientific innovation.

The topic of scientific development has been studied extensively in the philosophy of science and cognitive science. At a macro level, philosophers have contributed theories to the development of the broad scientific field. Representative work includes Popper's falsificationism (Popper, 1959), Fayerabend's incommensurability thesis (Fayerabend, 1962) and Kuhn's theory of scientific revolution (Kuhn, 1962). For instance, Kuhn characterized science as consisting of two alternating modes, "normal science" and "revolutionary science" (Kuhn, 1962). Whereas the periods of normal science are incremental and cumulative, the periods of revolutionary science involve a revision and reframing of pre-existing scientific ideas (Kuhn, 1962). At a micro level, scholars have suggested that the development of scientific concepts such as "temperature" or "H2O" often involves a gradual process of conceptual change (Carey, 2009; Chang, 2004, 2012).

Recent computational work has added to this line of inquiry by modeling topic trending in science (Griffiths & Steyvers, 2004; Prabhakaran, Hamilton, McFarland, & Jurafsky, 2016), the evolution of citation frames and networks (Jurgens, Kumar, Hoover, McFarland, & Jurafsky, 2018; Zeng, Shen, & Zhou, 2019), and the propagation of innovation in a specific scientific field (Jurafsky, 2015).

Our focus here differs from the existing array of work. We explore whether there is regularity in a scientist's innovative research outputs as they emerge over time. Although scientists vary widely in the subjects they study, we hypothesize how scientists conceive novel ideas over time should exhibit shared tendencies that reflect basic principles of human induction. There is some work from cognitive psychology suggesting that individual scientists undergo a process of conceptual change (Carey, 2009), and we extend this idea by probing incremental mechanisms that underlie the process of innovation in scientists.

Spatiotemporal mapping and chaining of scientific outputs

We describe two ideas that are central to the formulation of our framework: spatiotemporal mapping and chaining of scientific outputs.

Spatiotemporal mapping. We represent the main scientific outputs (i.e., published papers) throughout a scientist's career as dots on a spatiotemporal map. We use spatial proximity between dots to approximate semantic similarity of publications, and we timestamp each paper by its year of publication. We then infer a plausible innovation path for a scientist in question, by revealing one dot at a time as a publication emerges, and we ask whether the emerging order of publications into the future (at time $t+1$) is predictable given the spatiotemporal profile of existing publications (at time t). Figure 1 illustrates the map of scientific outputs from the Turing Award winner Geoffrey Hinton. What this map reveals is a compact summary of Hinton's scientific outputs over a 40-year period. It is clear that they have shifted from the initial focus on visual perception to the later focus on deep learning. What this map does not reveal directly is the hidden processes that underlie the shift: how one's earlier research ideas might influence the development of future ideas.

Chaining. We next explore plausible mechanisms that explain the emergence of future scientific outputs from existing ones, and we do so by constructing a connected path among the dots (or papers) over historical times. Our basic premise

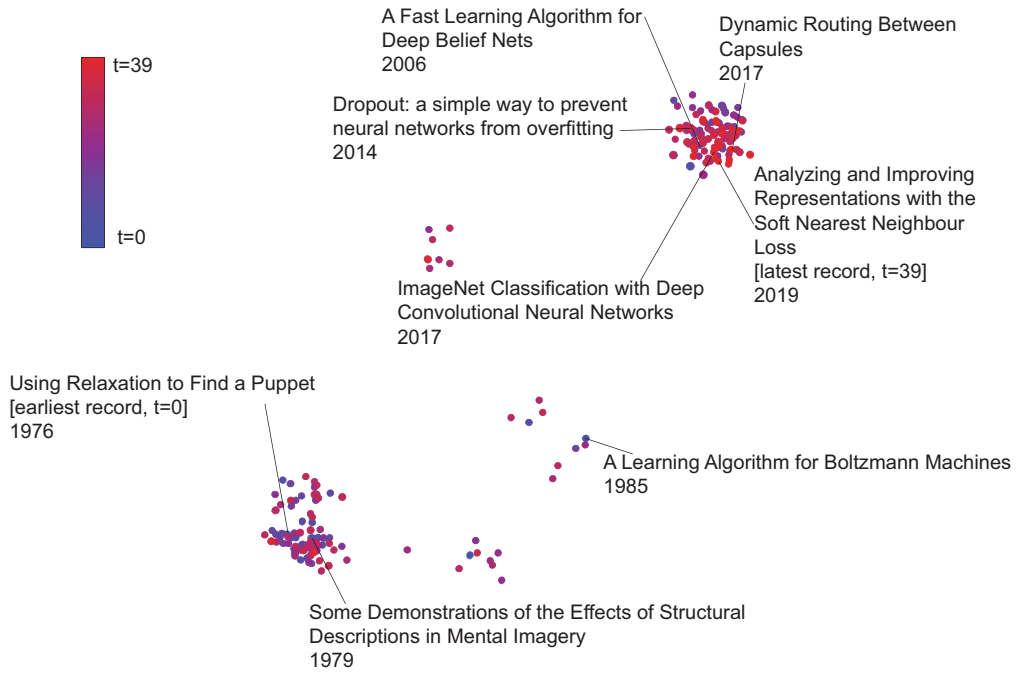

Figure 1: Illustration of Geoffrey Hinton's publications (1976-2019). Each dot represents a published paper projected onto a 2-dimensional space via tSNE (Maaten & Hinton, 2008). Year of publication is expressed in a cool-warm color gradient (warmness indicates recency). Distance between dots reflects semantic similarity between two papers in their Abstract embeddings.

is that the emerging order of publications is non-arbitrary, and certain ways of scientific innovation may be preferred by over others—statistically speaking among the scientists, because they are more cognitively natural or efficient.

Our hypothesis about the scientific innovation mechanisms is grounded in chaining, a prominent proposal on the growth of linguistic meaning. Chaining refers to a process that links novel ideas with existing ideas that are closely related, hence forming a chain-like structure over time (Lakoff, 1987; Bybee, Perkins, & Pagliuca, 1994; Malt, Sloman, Gennari, Shi, & Wang, 1999). Previous work has formulated chaining as probabilistic graph-traversal algorithms (Sloman, Malt, & Fridman, 2001; Xu, Regier, & Malt, 2016; Ramiro, Srinivasan, Malt, & Xu, 2018; Habibi, Kemp, & Xu, in press; Grewal & Xu, in press). More recently, it has been suggested that chaining can be best operationalized as an exemplar model, initially proposed as a theory of categorization—sometimes also known as the General Context Model (Nosofsky, 1986) and later extended to predicting the historical extensions of linguistic meaning in domains including numeral classifiers (Habibi et al., in press) and adjectives (Grewal & Xu, in press). Here we explore the connection between these computational chaining algorithms and the process of scientific innovation: we believe that the process of conceiving novel scientific ideas may resemble the process of linguistic innovation, and we test this proposal against historical publication records of scientists. Independent work from mathematical modelling has proposed that similar processes might underlie the emergence of novelties (Loreto, Servedio, Strogatz, & Tria, 2016), but to our knowledge there has been no formalization or comprehensive evaluation of chaining in scientists' innovation.

There is no strong reason *a priori* for scientist's publication order to be predictable, and in fact, many factors involved in the process of innovation that might forbid one from reversing the path (e.g., scientific collaboration, a point that we return to later). There is also no direct ground truth for evaluating our proposal on chaining, because only the scientists themselves would have gone through the actual thought processes. As we shall describe, we evaluate a set of models on their ability to predict the historical emerging order of papers of individual scientists, similar to how previous work predicts the emergence of word meaning (Ramiro et al., 2018). If the chaining models can reconstruct the emerging order of papers for different scientists, it will provide support for our hypothesis and reveal basic principles in the process of scientific innovation.

Computational methodology

Problem formulation. We formulate the problem as a temporal prediction problem. Given the set of papers S_t a scientist has published by time t, we want to predict the probability of a yet-to-emerge paper x^* from the same scientist to be ac-

tually published at the next available time step $t+1$. Our analysis treats each year as an individual timestep, so papers published within the same year are considered to emerge at the same time. Each model we specify below predicts which papers are most likely to emerge at the next timestep based on Luce's choice rule (Luce, 1959):

$$p(x^*|S_t) \sim \frac{f(x^*,S_t)}{\Sigma_{x\in S_t^*} f(x,S_t)} \quad (1)$$

where S_t^* is the set of papers yet to emerge (or be published) beyond t, and $f()$ specifies the particular model class. As a model predicts through time, S_t is updated with the ground truth papers at each timestep in order to minimize error propagation. Whenever "similarity" is mentioned below, we refer to the exponentiated squared Euclidean distance between paper embeddings x_1 and x_2 in semantic space.

$$sim(x_1,x_2) = exp(-d(x_1,x_2)^2) \quad (2)$$

Semantic representation. We capture the gist of a scientific paper by its content in the Abstract section. Specifically, we represent the semantics of each paper by a sentence embedding of the entire abstract generated by a pretrained BERT model (Reimers & Gurevych, 2019).

Computational models of chaining. We consider five main model classes adapted from recent work on semantic chaining (Ramiro et al., 2018; Habibi et al., in press; Grewal & Xu, in press). Each model class postulates a different mechanism of chaining that yields potentially different predictions in the emerging order of papers, and several of these models are related to neighbourhood-based chaining modelled closely by the exemplar theory (Nosofsky, 1986). Table 1 summarizes the model specifications of the $f()$ function in Equation 1.

1) *k-Nearest-neighbor models.* k-Nearest-neighbor (kNN) models are commonly used for classification and regression, but in this case we use them to predict which paper from S_t^* is likely to emerge next based on that paper's similarity to the k most similar papers that have already appeared in S_t. We tested values of k from 1 to 5. All of these models are slight variants of the exemplar model with hard and pre-specified neighbourhood sizes.

2) *Prototype model.* The prototype model is based on Rosch's work in categorization (Rosch, 1975). The prototype of a scientist's papers at S_t is taken to be the average of paper embeddings at S_t, and the probability of a paper emerging at time $t+1$ is proportional to its similarity to the prototype at time t. This results in a moving average against which new papers are compared.

3) *Progenitor model.* The progenitor model is a static variant of the prototype model in which the prototype is fixed at the first paper to emerge. The probability of a paper emerging at any timestep is then proportional to its similarity only to the first paper that emerges. This model suggests that a scientist's paper order is fully predictable given how related that paper is to the first paper a scientist has published, which means that the scientist might have progressed consistently in a restricted topic.

4) *Exemplar model.* The exemplar model is based on work by Medin and Schaffer, and Nosofsky (Medin & Schaffer, 1978; Nosofsky, 1986). In the exemplar model, the probability of a paper emerging at time t is proportional to its similarity to all other papers S_t. The most prominent exemplar model is the Generalized Context Model (Nosofsky, 1986), in which a sensitivity parameter s controls how sharply similarity tapers with increased (Euclidean) distance. We also consider a parameter-free version of this model assuming $s = 1$. The exemplar model can be viewed as a soft extension of the kNN models, in which k is equal to the number of total papers at each timestep but each paper is weighted differentially toward prediction.

5) *Local model.* The local model is a variant of the 1NN model wherein candidate papers are only compared against papers that emerged specifically at the previous timestep, to capture some recency effect (e.g., a paper emerging next is likely to be most related to the paper that just appears before it). We predict a paper's probability of arising at time t as proportional to its similarity to papers in $S_{t-1} \setminus S_{t-2}$, where S_i denotes the set of all papers that have emerged at time i or earlier.

Table 1: Specification of selected models with respect to the probability of paper x^* appearing at time t, given a set of existing papers at time t, S_t. $sim()$ is defined in Equation 2, and d refers to Euclidean distance.

Model	$f(x^*\|S_t)$
1NN	$max_{x\in S_t} sim(x^*,x)$
Prototype	$sim(x^*, prototype(S_t))$
Progenitor	$sim(x^*,x_0)$
Exemplar	$\Sigma_{x\in S_t} exp(-sd(x^*,x)^2)$
Local	$sim(x^*, S_{t-1} \setminus S_{t-2})$

Model scoring. Given the ground-truth sequence of emergent papers (from the publication dates), we score each model by its predictive probability assigned to the true sequence of papers. A better model should assign a higher predictive probability to the true sequence. We calculate the probability of any sequence $R = r_1, ..., r_N$ as:

$$p(R) = p(r_1) \times p(r_2|S_1) \times ... \times p(r_N|S_{k-1}) \quad (3)$$

where there are k timesteps in total. Note that $p(r_1) = 1$ since the starting point is fixed. It is also possible for multiple papers to arise at the first timestep, in which case all of their probabilities are set to 1. Denote n_k as the number of papers that specifically emerge at the k-th timestep.

We compare our models against a random guess baseline

postulating any sequence is as likely as any other:

$$p(R_{rand}) = 1 \times \frac{1}{N - n_1 + 1} \times ... \times \frac{1}{N - n_k + 1} \quad (4)$$

For each non-random model, we define the probability of $p(r^*|S_t)$ through Equation 1 as specified. For each paper that appeared at a timestep, we evaluated its probability with respect to only papers that emerged at a different timestep and itself (since several papers can appear at a single time point).

Data collection and processing

In this initial exploration, we focused on analyzing scientific innovation in computer science researchers, although our framework is applicable to the analysis of other scientific fields. The cultural norm of publishing in computer science is primarily conference-based (Vrettas & Sanderson, 2015), so it provides us with useful time-locked publication data and a relatively high number of publications per scientist.

More specifically, we chose to work with 70 Turing Award winners (Association for Computing Machinery, 2019) because 1) their work has been highly recognized and is representative of the field 2) their academic profiles represent a decent temporal span of researchers from 1966 to 2018. While this is not a random sample of scientists, we would expect any regularity displayed in the emergence of innovation in the general scientific population to also be present in this sample.

Here we are not particularly concerned with the impact of a specific scientific paper, but rather its semantic relationships with other papers by the same individual. Nevertheless, it is possible that the type of work that scientists do, as well as individual differences between scientists would affect the models that best characterize them, and that this differs between a randomly selected group and a widely recognized group. This is not within the scope of this paper, but is a potential direction for future work.

Year, title, co-authors, and abstract data for the list of Turing Award winners were extracted from DBLP Computer Science Bibliography (`https://dblp.uni-trier.de/`) which is a standard database for bibliography in computer science. In order to restrict papers to scientific research, only journal articles and conference or workshop papers were retained.

Miscellaneous publications (opinion pieces, obituaries, panel discussions, interviews) were removed when it was clear that a publication fell into one of these categories. Words were lemmatized, and functional words, punctuation, bracketed words, and markup were removed in the calculation of average embeddings.

Results and evaluation

We focus on reporting two main aspects: 1) whether there is evidence for chaining in the emergence of scientific papers from different scientists; 2) how degree of collaboration might affect chaining models in predicting scientists' innovation trajectories.

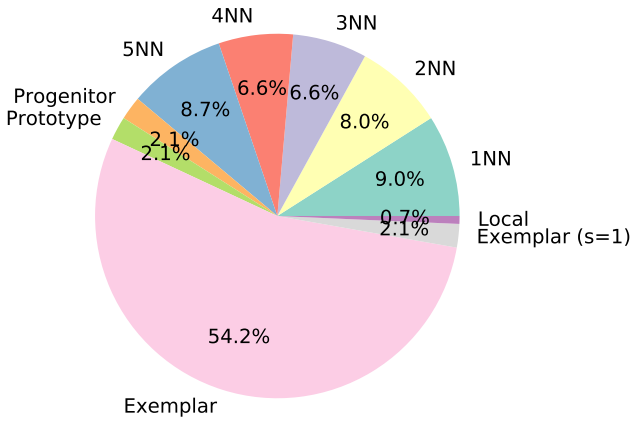

Figure 2: Summary of winner-take-all predictive analysis of all models based on data from 55 Turing Award winners.

Predominance of the exemplar model. To find the best-performing value of the sensitivity parameter s in the exemplar model, we performed a grid search over the range [0, 100]. To avoid giving this model an advantage over the other models that are parameter free, we used held-out data with a 20/80 train-test random split among the 70 scientists to evaluate all of the models. A larger test set was used in the interest of preserving as many scientists as possible for our analysis (55 scientists were left in the test set). The best s value was found to be 0.02.

We ran each class of models on data from the 55 Turing prize winners on an individual basis. To evaluate which of the models best accounted for scientists' extension of work, we performed a winner-take-all comparison of the models based on log predictive probability. If two or more models did equally well in predicting the paper sequence, they would receive equal credit, though each scientist's best model(s) received equal weight in the final tally.

The best-performing models are summarized in Figure 2. The overall best model is the exemplar model with $s = 0.02$, explaining 54.2% of the scientists that dominates the remaining models. A residual 38.9% of scientists are best explained by one of the kNN models, while progenitor and prototype models account for only 2.2% of scientists and the local model accounts for 0.7% of scientists.

We also visualized the average log-likelihood ratio against the random model for each model over the 55 Turing prize winners, and the results appear in Figure 3. These results align with the winner-take-all results, with the exemplar model achieving the highest average log-likelihood ratio, and the progenitor, prototype, and local models achieving the lowest average log-likelihood ratios. There is generally more variation among individual scientists and less variation within models for an individual scientist.

All kNN models are very closely correlated (by construction), such that a given scientist would tend to either be best

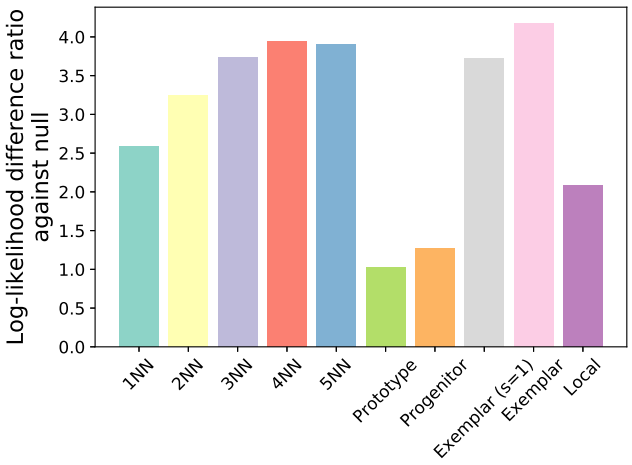

Figure 3: Summary of average log-likelihood ratio over the random model of all models based on data from 55 Turing Award winners.

explained by the exemplar model or be best explained by several of the kNN models. Figure 4 confirms this is the case by showing the inter-model correlations in prediction across the 70 data points. The class of kNN models correlates most highly with one another. The local model and 1NN are highly correlated with Pearson $\rho = 0.991$. The progenitor model does not correlate as much with any of the other models, for instance with 1NN ($\rho = 0.684$). This corroborates the account that there is more variation between scientists than between model performance on a single scientist.

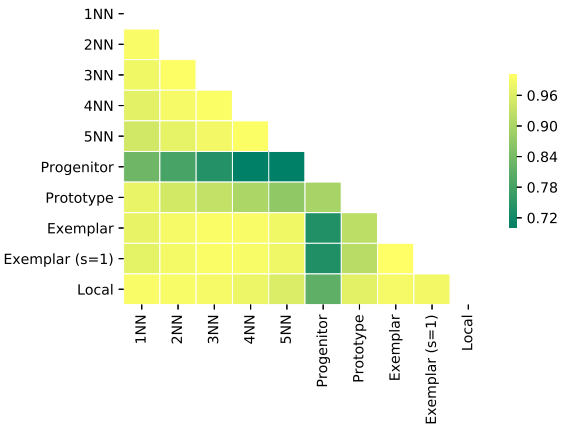

Figure 4: Inter-model correlations in prediction.

Overall, this initial set of results provides support for the idea that how scientists innovate may involve a mechanism similar to that captured by a weighted exemplar model, more so than prototype-based models or nearest-neighbour models. These results resonate with findings made in the linguistic domains of historical adjective extension and numeral classifiers (Grewal & Xu, in press; Habibi et al., in press).

Predictability in scientific trajectories and collaboration. Our analyses so far have focused on comparing different model sets but not how these models fare with random drift, which is plausible for scientific innovation.

To address this issue we analyzed each scientist individually. In each case, the probability of the ground-truth sequence of papers occurring in each model was compared to the probability of the sequence occurring in the random model, which is determined by Equation 4. Figure 5 shows that most models perform above chance for most scientists, and at least one model performs above chance for 87.3% of scientists (48/55).

One pertinent question is to what extent these results are affected by collaboration. Scientists sometimes decide what to work on based on what their collaborators are working on, so the process of extension is not purely one person's cognitive process. To examine the effect of collaboration on these results, we ran the models on papers written by a low number of co-authors. We also considered a separate run where only papers written by a scientist as the leading or final (senior) author were included.

The results are summarized in Table 2. Model performance above chance peaks when we only considered lead-author papers, but there is not a significant difference between the results for lead-author papers and the results with no authorship restrictions in place. This indicates that there may not be a significant effect of authorship on the ability for models to predict the sequence of papers above chance.

Table 2: Number of scientists for which models performed above chance, evaluated via log-likelihood ratio against random of the ground-truth sequence.

Paper authorship	Number (%) of scientists predicted above chance
Lead author	**50/55 (90.9%)**
1 author	49/55 (89.1%)
2 authors	47/55 (85.4%)
3 authors	49/55 (89.1%)
No authorship restriction	48/55 (87.3%)

To verify whether the exemplar model persists to be the best explanatory model under these different authorship conditions, we repeated the winner-take-all analysis and summarize the breakdown of model performance when models predicted above random in Table 3.

In each case, we confirm our initial finding that the exemplar model is the best overall in comparison to the other alternative models. We note that there are slight variations across conditions, with the 1NN and 2NN models accounting for a greater proportion of scientists in the 1 author condition compared to baseline, and the exemplar model accounting for a lesser proportion in the 1 author condition.

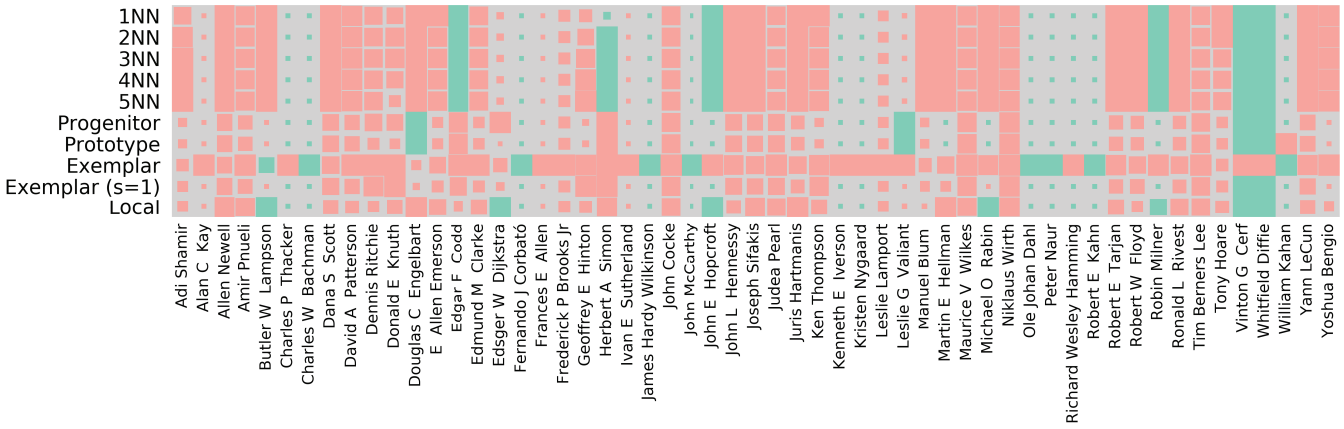

Figure 5: Model performance above random on 55 Turing Award winners (1966-2018), with no authorship restriction. Green squares indicate that the model performed below chance, while red squares indicate that the model performed above chance. The size of squares is proportional to the absolute value of the log-likelihood ratio between the model's assigned probability and the random probability of a true sequence, relative to either the best ratio for that scientist (if the ratio is above 1) or the worst ratio for that scientist (if the ratio is below 1).

Table 3: Proportion of scientists predicted above chance that are best explained by each model under different authorship conditions.

Paper authorship	1NN	2NN	3NN	4NN	5NN	Progenitor	Prototype	Exemplar	Exemplar (s=1)	Local
Lead author	7.1%	7.6%	6.2%	6.2%	8.3%	2.3%	0.2%	**59.2%**	2.3%	0.6%
1 author	15.6%	18.7%	9.0%	2.7%	4.8%	1.3%	0.2%	**33.5%**	10.6%	3.7%
2 authors	7.2%	7.2%	5.7%	5.7%	7.9%	2.9%	0.7%	**58.7%**	2.9%	1.1%
3 authors	8.3%	4.7%	5.4%	5.4%	7.4%	2.5%	0.5%	**62.5%**	2.5%	0.8%
No restriction	9.0%	8.0%	6.6%	6.6%	8.7%	2.1%	2.1%	**54.2%**	2.1%	0.7%

Table 4 highlights some authors for which either the Exemplar or the 1NN model performed especially well, as indicated by the gap in log-likelihood ratio between that model and the next best model on a particular scientist. One question that we investigated was whether or not the performance of these models correlates with the number of papers that a scientist has published, but we found that there is a small correlation between number of papers published and the comparative advantage of that model over others (1NN: $\rho = 0.211$, Exemplar: $\rho = 0.126$), but neither is statistically significant (1NN: $p = 0.122$, Exemplar: $p = 0.358$).

Table 4: Top 5 scientists for which the Exemplar and 1NN models performed the best.

Exemplar	1NN
Edmund M. Clarke	E. Allen Emerson
Robin Milner	Ken Thompson
Leslie Lamport	Allen Newell
Judea Pearl	Butler W. Lampson
John E. Hopcroft	Charles W. Bachman

Conclusion

We present to our knowledge the first computational account for the processes of scientific innovation in individual scientists. None of the models we considered is complex, but each model is grounded in the existing literature and easily interpretable. The exemplar model performs better than the alternative models in characterizing well-known scientists' paper sequences over time in the field of computer science.

Our results provide support for the view that the processes underlying how scientists develop ideas over time may rely on cognitive mechanisms of chaining that closely resemble those found in linguistic innovation (Habibi et al., in press; Grewal & Xu, in press), but variation exists among different scientists. Future work should assess the generality of these findings in scientific domains other than computer science.

Acknowledgments

We thank Geoffrey Hinton for helpful suggestions. We also thank Muhammad Ali Khalidi for pointers and discussion. YX is funded through an NSERC Discovery Grant, a SSHRC Insight Grant, and a Connaught New Researcher Award.

References

Association for Computing Machinery. (2019). *Chronological listing of a.m. turing award winners.* Retrieved 2019-01-28, from https://amturing.acm.org/byyear.cfm

Bybee, J. L., Perkins, R. D., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. Oxford University Press.

Chang, H. (2012). *Is water H2O?: Evidence, realism and pluralism*. Springer Science & Business Media.

Fayerabend, P. (1962). Explanation, reduction, and empiricism. In H. Feigl & G. Maxwell (Eds.), *Scientific explanation, space, and time.* University of Minneapolis Press.

Grewal, K., & Xu, Y. (in press). Chaining and historical adjective extension. In *Proceedings of the 42nd annual meeting of the cognitive science society.*

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, *101*(suppl 1), 5228–5235.

Habibi, A., Kemp, C., & Xu, Y. (in press). Chaining and the growth of linguistic categories. *Cognition*.

Jurafsky, D. (2015). *The spread of innovation in speech and language processing.* Retrieved from https://www.icsi.berkeley.edu/icsi/morganfest (ICSI Morganfest)

Jurgens, D., Kumar, S., Hoover, R., McFarland, D., & Jurafsky, D. (2018). Measuring the evolution of a scientific field through citation frames. *Transactions of the Association for Computational Linguistics*, *6*, 391-406.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Loreto, V., Servedio, V. D. P., Strogatz, S. H., & Tria, F. (2016). Dynamics on expanding spaces: Modeling the emergence of novelties. *Creativity and Universality in Language*, 59–83.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. Wiley, New York.

Maaten, L. v. d., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, *40*, 230–262.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Popper, K. (1959). *The logic of scientific discovery*. Basic Books.

Prabhakaran, V., Hamilton, W. L., McFarland, D., & Jurafsky, D. (2016). Predicting the rise and fall of scientific topics from trends in their rhetorical framing. In *Proceedings of the 54th annual meeting of the association for computational linguistics.*

Ramiro, C., Srinivasan, M., Malt, B., & Xu, Y. (2018). Algorithms in the historical emergence of word senses. *Proceedings of the National Academy of Sciences*, *115*, 2323-2328.

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proccedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing.*

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*, 192–233.

Sloman, S. A., Malt, B. C., & Fridman, A. (2001). Categorization versus similarity: The case of container names. In U. Hahn & M. Ramscar (Eds.), *Similarity and categorization* (pp. 73–86). New York: Oxford University Press.

Vrettas, G., & Sanderson, M. (2015). Conferences versus journals in computer science: Conferences vs. journals in computer science. *Journal of the Association for Information Science and Technology*, *66*, 2674–2684.

Xu, Y., Regier, T., & Malt, B. C. (2016). Historical semantic chaining and efficient communication: The case of container names. *Cognitive Science*, *40*, 2081–2094.

Zeng, A., Shen, Z., & Zhou, J. e. a. (2019). Increasing trend of scientists to switch between topics. *Nature Communications*, *10*, 3439.