Title

A Picture is Worth 7.17 Words: Learning Categories from Examples andDefinitions

Permalink

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors

Moskvichev, Arseny
Tikhonov, Roman
Steyvers, Mark

Publication Date

2019

A Picture is Worth 7.17 Words: Learning Categories from Examples and Definitions

Arseny Moskvichev (amoskvic@uci.edu)
University of California, Irvine

Roman Tikhonov (r.tikhonov@spbu.ru)
Saint Petersburg University, St. Petersburg, Russia
National Research University Higher School of Economics, St. Petersburg, Russia

Mark Steyvers (mark.steyvers@uci.edu)
University of California, Irvine

Abstract

Both examples and verbal explanations play an important role in learning new concepts and categories. At the same time, learning from verbal explanations is not accounted for in most category learning models, and is not studied in the traditional category learning paradigm. We propose a *rational category communication* model that formally describes the process of communicating a category structure using both verbal explanations and visual examples in a pedagogical setting. We build our model based on the assumption that verbal instructions are best suited for communication of crude constraints on a category structure, while exemplars complement it by providing means for finer adjustments. Our empirical study demonstrates that verbal communication is indeed more robust to changes in stimuli dimensionality, but that its efficiency is adversely affected when distinguishing between categories requires perceptual precision. Communicating through examples has a reversed pattern. We hope that both the proposed experimental paradigm and the computational model would facilitate further research into the relative roles of verbal and exemplar communication in category learning.

Keywords: categorization; category learning; computational modelling; communication efficiency; communication channels

Introduction

Humans have a variety of information sources available to enrich or expand their knowledge. Imagine a person encountering an unfamiliar word or concept. She may infer its meaning from examples of how it is used, consult a dictionary, or use a combination of examples and definitions to understand a word or concept. In many cases, any of these sources alone is not sufficient (Fischer, 1994; Nagy, Herman, & Anderson, 1985).

Similarly, multiple sources of information are also often used to communicate a category or a concept. Imagine a family forest trip where a parent wants to teach their child about poisonous mushrooms. It is easy to envision a parent instructing their child through definitions, e.g., not to collect pale, thin-legged mushrooms with a flat cap since they are usually poisonous. Is also easy to imagine this parent giving examples, e.g. “look: this is one of the poisonous mushrooms I told you about”. A key difference is that the former involves a verbal explanation of a rule, while the latter relies on non-verbal ways of concept communication (relevant examples only need to be pointed at). Contrary to the situation with word learning, however, in the context of perceptual categories, the relative contributions of verbal- and example-based communication are not well understood.

We know, however, that example- and verbal-based communication are not redundant: different aspects of category and concept knowledge may require different means of communication. Verbal instructions are well suited for communication of abstract rules, but give little information about specific stimuli characteristics (Longman, Milton, Wills, & Verbruggen, 2018). Examples, in turn, provide contextual information and help to understand how to apply knowledge to a particular problem (Reed & Bolstad, 1991; Fischer, 1994). Thus both example- and verbal-based communication play a significant role in shaping human learning. As such, they should be incorporated into contemporary theories and computational models of category acquisition.

In this work, we focus on the question of what are the fundamental differences between the verbal and exemplar channels of communication. We formalize the aforementioned intuitions about these differences and propose a computational model of the process. We also run an empirical study that investigates how people communicate perceptual categories using different combinations of communication channels. In particular, we investigate how different characteristics of a category structure affect the efficiency of verbal- and exemplar- based category communication.

Related work

The problem of communicating knowledge spans a broad range of disciplines, including educational and cognitive psychology, logic, linguistics, mathematics, and philosophy.

In the area of machine learning, there is a range of works on the problem of knowledge communication (e.g., (Winston, Binford, Katz, & Lowry, 1983)). In particular, there is a growing interest in the problems of few- and zero- shot learning techniques that focuses on learning through language without ever seeing an example (DeJong & Mooney, 1986). Notably, Mitchell (Mitchell, Keller, & Kedar-Cabelli, 1986) looked specifically into the ways of learning artificial categories from verbal explanations. In most cases, these attempts are, however, centered on applications in their respective domains and do not aim understand or model the fundamental roles that different communication systems play in human interaction and learning.

Surprisingly, verbal communication has not received much attention in empirical studies of category learning and has been largely ignored in corresponding computational models. Well-established paradigms for category learning focus on the communication and acquisition of categories through examples only and miss one of the critical sources of information used in real-world situations. Considering the overwhelmingly important role of verbal communication in education and the impact of internal verbalization on the learning outcomes (Vinner, 2002; Lombrozo, 2012; Williams & Lombrozo, 2010, 2013), this omission makes the well known ironic definition of category learning as the "class of behavioral data generated by experiments that ostensibly study categorization" (Kruschke, 2008) exceedingly appropriate.

We see two related reasons for this apparent oversight. First, the fact that people use definitions to acquire knowledge is so apparent, and, at the same time, so difficult to model rigorously, that it is very tempting to ignore either as "boring" or "impractical" to study. It is sometimes seen as an unstated assumption that verbal communication would allow to simply transfer the category knowledge.

Second, learning from definitions is inherently pedagogical, and, until recently, we lacked the tools to model such situations. Historically, category learning literature focused on extracting knowledge from a neutral environment (although there are notable exceptions: (Avrahami et al., 1997)), and the formal apparatus for modeling pedagogical reasoning in category learning was developed only recently (Shafto, Goodman, & Griffiths, 2014; Aboody, Velez-Ginorio, Laurie, Santos, & Jara-Ettinger, 2018; Frank & Goodman, 2012).

Even though recent years have witnessed a revived interest in empirical studies of these distinct ways of learning (Liefooghe, Braem, & Meiran, 2018; Longman et al., 2018), the modeling aspect is critically lacking.

Overall, we believe that now, when we have the tools to model pedagogical reasoning in category learning setting, it is a good time to make a step towards a formal model of both explanation- and example-based category learning.

Relation to categorization models

While the attempts to introduce learning based on verbal explanations into category learning models are scarce, many of the prominent categorization models could be naturally extended to partially account for verbal communication. For example, in the ALCOVE model (Kruschke, 1992), verbal communication could be introduced as transferring attention weights, thus speeding up subsequent example-based learning. On the other hand, there is no clear way to introduce purely verbal communication into this or most of the other exemplar models.

In the case of RulEx (rules with exceptions) model (Nosofsky, Palmeri, & McKinley, 1994), verbal communication could be introduced as a direct rule transfer, while examples may serve as illustrating exceptions, or as a way of adjusting rule boundaries.

Another prominent categorization model, COVIS (Ashby, Paul, & Maddox, 2011), includes the verbal (rule-based) and procedural (information-integration) components. These names partially acknowledge the potential importance of verbal reasoning, and difference in learning dynamics for "verbalizable" and "non-verbalazable" categories were extensively studied by G. Ashby (Ashby et al., 2011). At the same time, the verbal system is mostly seen as a component of internal learning dynamics, and its relation to knowledge communication is not usually studied.

Overall, there are many potential ways to introduce verbal communication into existing categorization models. At the same time, learning from verbal explanations is inherently pedagogical (somebody has to produce the explanations for a student). Therefore, we find it most promising to approach the problem from the rational analysis perspective which already offers an elegant account of pedagogical reasoning in category learning. In the next sections, we describe our approach.

Computational Model

We build upon the rational account of pedagogical reasoning, introduced in (Shafto et al., 2014). That work provided an answer to the question of how a rational teacher should select the most useful example to help a rational student learn a specific category.

In their approach, a rational teacher aims to choose an example that would maximize the student's learning outcome (probability of selecting a correct hypothesis). Thus, the teacher needs a model of the student. A rational student will also try to understand why their teacher selected a specific example which means that a student has to model the teacher. The authors formalize it as a pair of equations:

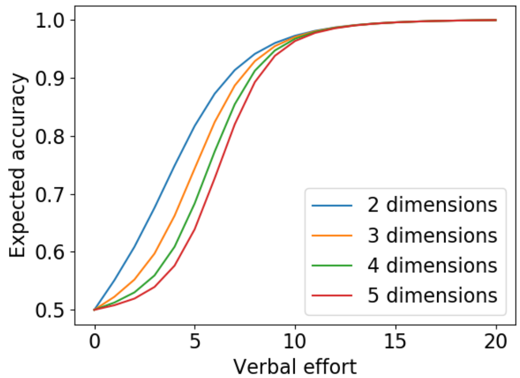
$$P_{teacher}(d|h) \propto (P_{learner}(h|d))^{\alpha} \quad (1)$$

and

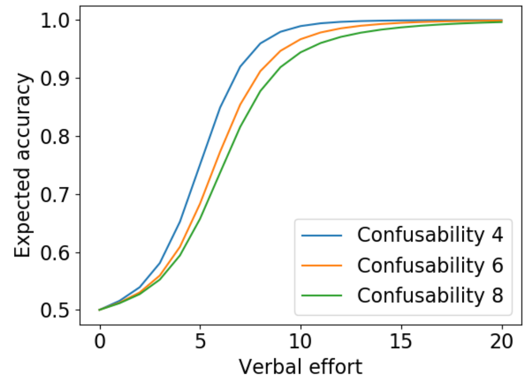
$$P_{learner}(h|d) = \frac{P_{teacher}(d|h)P(h)}{\sum_{h'} P_{teacher}(d|h')P(h')} \quad (2)$$

Where h stands for the hypothesis and d stands for the data.

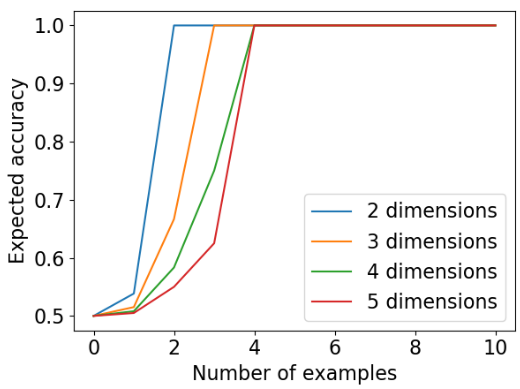
Equation 1 states that the teacher should select data points proportionally to the posterior probability of the correct hypothesis that a learner would infer after seeing these examples. Parameter α reflects how much is the teacher inclined to sample the most informative example. Thus $\alpha = 1$ corresponds to probability matching, while $\alpha = \infty$ corresponds to a deterministic selection of the best example. In the original model, an α of 1 was used in all experiments (Shafto et al., 2014).



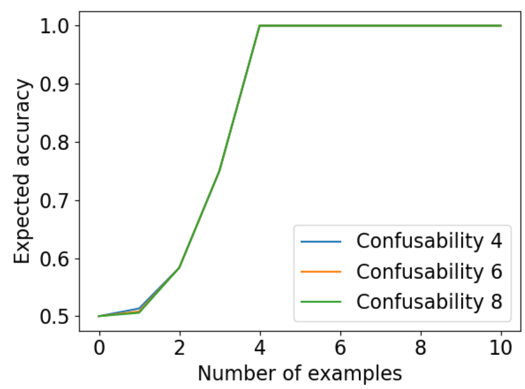
(a) The curves converge early.



(b) The curves only converge near their asymptote.



(c)



(d)

Figure 1: Simulation results of the effect of stimuli dimensionality and perceptual confusability on expected accuracy. In particular, the efficiency of exemplar channel of communication is not affected by perceptual confusability, while the number of dimensions has a noticeable impact on it. Additionally, verbal communication curves quickly converge for different stimuli dimensionality. Thus, the number of dimensions matters for low-quality verbal explanations, but their impact fades as the quality of explanations increases. On the other hand, perceptual confusability continues to matter even for high-quality explanations, highlighting the intuition that small perceptual differences may be very difficult to verbalize.

Equation 2 states that the learner should select hypotheses proportionally to how likely a rational teacher is to generate the available data under these hypotheses. A solution could be obtained by substituting one into another and iteratively updating some initial estimate until convergence.

In order to incorporate verbal communication into this model, as well as to make the model more broadly applicable, we need to make a number of changes. In the next sections we will first describe them conceptually, and then write down the resulting equations.

Sequential sampling

In (Shafto et al., 2014), authors exhaustively enumerated all possible datasets that could be communicated. Thus, if a teacher wants to show a student three examples, choosing among N possible examples every time, the space of possible datapoints is going to be N^3. This exponential data space is very limiting even for simple category learning tasks if a training session consists of more than just a few examples.

We assumed that the data is selected sequentially, in a greedy fashion. Thus, if a teacher had to select three examples, she would first select a single example that maximizes the probability of the correct category, then selects the second example conditional on the event that the student already saw the first one, and so on.

This does not guarantee an optimal sample in general, but it makes the model applicable in realistic conditions. For example, in traditional category learning experiments, which often include a large number of trials and high-dimensional stimuli as well as sequential, interactive teaching. In principle, it is also possible to combine sequential sampling with exhaustive enumeration, by adding a tractable number of examples on each step.

Formally, we rewrite Equations 1 and 2 introduce sequential dependencies (an addition to recursive teacher-student dependencies already present).

$$P_{teacher}(d_i|d_{i-1},...,d_1,h) \propto (P_{learner}(h|d_i,d_{i-1},...,d_1,h))^{\alpha} \quad (3)$$

$$P_{learner}(h|d_i...d_1,h) = = \frac{P_{teacher}(d_i|d_{i-1},...,d_1,h)P_{learner}(h|d_{i-1},...,d_1)}{\sum_{h'} P_{teacher}(d_i|d_{i-1},...,d_1,h')P_{learner}(h'|d_{i-1},...,d_1)} \quad (4)$$

Where d_i is a data point selected by a teacher on step i. This completes the formal description of the model for the case when all d_i are examples.

Verbal communication

The key problem we have to solve is incorporating verbal communication into the model, i.e., handling the case when d_i is a verbal explanation.

Explicitly mapping language to category structures that are communicated is an extremely difficult task. We sidestep the issue by modeling the process at a higher level: we simply assume that verbal communication channel allows us to transfer the information about which hypothesis is correct. If we view the problem this way, the problem of selecting which category structure to communicate is not relevant: we could assume that the teacher always intends to communicate the correct hypothesis.

This channel of communication has its limitations, which may depend on the category structure. For example, some hypotheses could be difficult or impossible to formulate verbally (Ashby et al., 2011), and some information could be lost due to miscommunication or misunderstanding.

To account for these phenomena, we assume that the channel is noisy. That is, even though the teacher always intends to communicate the correct hypothesis and "sends" it through the verbal channel, due to noise, instead of receiving an unambiguously decoded hypothesis, a student only receives a sample from a distribution over all possible hypotheses. The shape of this distribution depends on the hypothesis being sent and is determined by the noise model.

Noise model

It is reasonable to assume that the noise corruption is more likely to turn a hypothesis into a similar hypothesis, as opposed to turning it into something entirely unrelated. There are, however, different ways to define this similarity metric for the corruption model.

One approach is to restrict oneself to a certain class of rules and then define similarity in some intuitive way. One option would be to rely on syntactic similarity between formal expressions defining a concept (this would be similar in spirit to (Goodman, Tenenbaum, Feldman, & Griffiths, 2008)), or in some other way manually define the distance function between any two hypotheses.

We want our model to be applicable in a wide range of categorization experiments, and thus we chose not to rely on any specific choice of the hypothesis space. Instead, we model similarity between two categories simply as the similarity in the pattern of their predictions. Thus, two categories (hypotheses) are the maximally similar if they predict the same answer for all examples, and they are maximally dissimilar if they always predict different answers. We find this definition highly neutral as it builds upon the most basic definition of equality of categories: the categories are the same if the sets of things that belong to these categories are equal.

Apart from being flexible and unopinionated, our noise model captures some fundamental and intuitive properties of language: its ability to transfer the gist of the situation in broad-brush terms, and its difficulty in exactly communicating perceptual experiences. Instead of being hard-coded into the model, these properties naturally emerge from the concept similarity definition that we employed.

For example, when two rule-based categories differ only slightly in the thresholds that define them, or if two prototype-based categories differ slightly in prototype means, there would likely only be a few examples that would be misclassified if we confuse two such concepts. Thus, these categories would be similar according to our definition, and it would be difficult to discriminate between them using the verbal channel of communication.

At the same time, if two rules differ in the dimensions that are considered relevant for it, or if some dimension is "reversed" - the ramifications of confusion between such two rules would be dramatic. Such rules would be very dissimilar according to our definition, and it would be easy to distinguish between them using the verbal channel of communication.

Verbal effort

There are good explanations and there are bad ones. The same concept could be explained clearly, leaving little or no uncertainty on the student's side, or it could leave the student confused, knowing little more than before.

In order to capture this intuition, we introduce a concept of *verbal effort*. The more *verbal effort* a teacher puts into her explanation the less uncertainty there is about what was the communicated category.

Putting it together

In order to fully specify the model, we start with the Equations 3 and 4, and complement them with the case when d_i is a verbal message via the Equation 5.

$$\begin{aligned} P(h|d_i,...,d_1) \propto P(d_i|d_{i-1},...,d_1,h)P(h|d_{i-1},...,d_1) = \\ = \left[\sum_{d_i^{sent}} P(d_i|d_i^{sent})P(d_i^{sent}|h)\right] P(h|d_{i-1},...,d_1) = \\ = P(d_i|d_h^{sent})P(h|d_{i-1},...,d_1) \quad (5) \end{aligned}$$

Where d_h^{sent} is the index of the correct hypothesis. The last equality holds since the teacher always (i.e. with probability

1) attempts to verbally communicate the true hypothesis.

Lastly, we define

$$P(d_i|d_i^{sent}) \propto \exp\{\sigma_{d_i,d_i^{sent}} \cdot \eta\} \tag{6}$$

Where $\sigma_{d_i,d_i^{sent}}$ is the correlation in predictions between the communicated hypothesis index d_i^{sent} and d_i, the (potentially noise corrupted) index of the received hypothesis. The softmax scale parameter $\eta \in [0,\infty)$ is the *verbal effort*. A verbal effort of zero corresponds to complete randomness: nothing useful was transmitted verbally. A verbal effort of infinity, in contrast, allows one to exactly identify the correct hypothesis. Currently, we fixed the steps on which the verbal communication occurs, but this restriction could be relaxed.

Overall, Equations 3, 4, 5, and 6 provide a formal definition of our model. See supplementary materials for the model implementation.

Evaluation

In the next sections we describe the experimental setting on which we collected both empirical and simulation data to test test the viability of our model.

Experiment

Method

Participants We recruited 357 participants (169 as *teachers* and 188 as *students*) through Amazon Mechanical Turk. They were native English speakers from the US. We excluded from the analysis teachers who did not reach predefined 85% accuracy threshold ($n = 40$) or failed to follow the instructions ($n = 28$), resulting in a final sample of 101 teachers.

Materials Schematic representations of fish (Rosedahl & Ashby, 2018) with possible variations in up to five visual features (fin, tail, belly color, etc.) were used as stimuli. We varied three independent variables between the participants: 1) *stimuli dimensionality* (two, three, or four dimensions) – the number of visual features varying in the presented stimuli, 2) *perceptual confusability* (low/high) – the visual similarity between stimuli of two categories, and 3) *rule type* (one- or two-dimensional). Exact visual features related to the rule dimensions were selected randomly.

Procedure Teachers learned the categorization rule by observing two sets of 15 stimuli labeled *Examples of type A* and *Examples of type B* (see Figure 2). Stimuli were presented simultaneously. Teachers had no time constraints and were able to explore each stimulus in more details by enlarging it. In the test phase, teachers had to categorize 30 stimuli presented sequentially (15 stimuli of each category including at least eight stimuli that were not presented before). Teachers who achieved the accuracy threshold of 85% in the test phase were asked to generate three training sets to teach other participants. There were three teaching formats (the order was counter-balanced across the teachers): *verbal*, *examples*, and *mixed*. In the *verbal* format teachers had to provide instructions that allow categorizing the stimuli. In the *examples* format they had to generate new stimuli of two different categories without any verbal explanations (category labels were provided). In the *mixed* format teachers were allowed to use both verbal instructions and visual examples (see Figure 2). Teachers could use as many words or visual examples as needed to explain the categorization rule, but they were instructed to be concise in their explanations and use only the minimum required amount of examples.

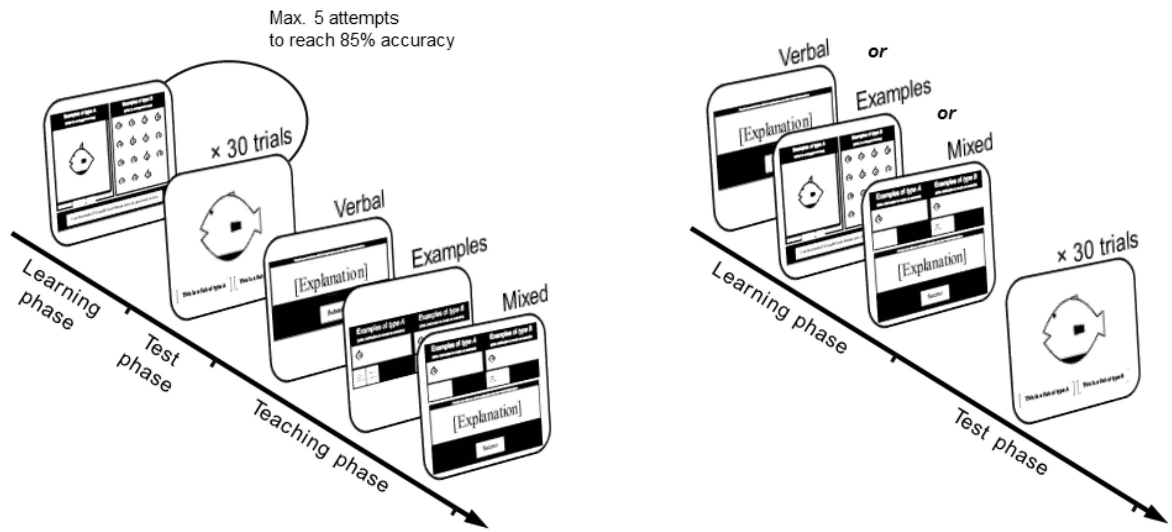
Students were randomly assigned to one of three learning conditions (verbal explanations, visual examples, or mixed), and received corresponding training materials prepared by one of the teachers. There were no time limits for the learning phase. The test phase was similar to the teachers' group.

Results

Students' performance More than 67 percent of students achieved 75% threshold criterion with median accuracy of 93 percent. Unfortunately, it results in overly low variability in the student accuracy variable. Some clear patterns were still present: one-dimensional rules result in higher performance (.85) than two-dimensional (.72), $p < .001$. As well as higher perceptual confusability decreased students' accuracy from .84 to .76 ($p < .001$). However, it would be impossible to capture the more subtle interaction effects that are relevant to our study. Initially, we planned to investigate the effects of text length and explanation numbers on the students' accuracies, but students' surprisingly good performance rendered this approach impractical. Thankfully, we could switch to another interpretation to still gain insight into the problem. Since the teachers were able to create learning materials that in most cases allowed students to master the concept, we could focus on the study materials themselves: did the teachers adjust their teaching strategies to the situation? We used a Poisson regression and Generalized Estimating Equations approach to account for the teacher-to-teacher individual differences. We applied robust variance estimation techniques to compensate for potential model misspecifications.

Words per picture The average number of the visual examples provided by teachers was 4.28 in the mixed condition and 4.86 in the examples condition. The average length of the explanations increased from 28.37 in the mixed condition to 34.86 in the verbal condition because of the absence of visual examples. That is, to achieve comparable performance, the teachers needed to write approximately $34.86 \div 4.86 = 7.17$ words per example. These values could be used to map the verbal effort variable used in the computational model to the number of words in the explanation and thus put it on a more intuitive scale.

Predictors of text length and number of examples We found statistically significant effects of the rule type ($\beta = .58, p < .001$), the perceptual confusability ($\beta = .22, p = .044$), and the presence of visual examples ($\beta = -.19, p = .004$) on the length of verbal explanations (in symbols). The effects of stimuli dimensionality were not statistically signifi-



(a) Teachers

(b) Students

Figure 2: Experimental procedure illustration

cant ($\beta = -.03, p < .655$). However, the number of visual examples was predicted only by the stimuli dimensionality ($\beta = .18, p = .012$) and the rule type ($\beta = .46, p = .025$). There were also marginally significant effects of the presence of verbal explanations ($\beta = -.12, p = .052$) and the interaction of the perceptual confusability and the stimuli dimensionality ($\beta = .15, p = .089$). The effects of perceptual confusability were not statistically significant ($\beta = -.29, p = .149$).

Simulation results

We used identical experimental settings to test the performance of our computational model. We obtained initial estimates of $P(d_i|h)$ using a strong sampling assumption, and then iteratively updated them until convergence.

As shown in Figure 1, the simulation results closely correspond to the patterns we observed in the experiment. It is important to mention, however, that the behaviour depicted on Figure 1d depends on the choice of the parameter α. We used $\alpha = 1.1$ in our experiments.

Apart from capturing the key dynamics present in our data, the model also makes a range of important predictions and provides rich opportunities for further experimentation. For example, it is able to capture the mutually enriching nature of verbal and exemplar communication channels. Thus, it is possible to model situations in which using verbal explanations and exemplars together leads to dramatic leaps in performance, allowing to reach maximum accuracy, while individual channel performance is mediocre at best (0.76 for exemplars, 0.53 for verbal communication).

Discussion and conclusion

We see the main impact of our paper in identifying a fundamental limitation characteristic of most existing human category learning models (little to no account for the verbal communication) and proposing a principled and broadly applicable model to account for these phenomena.

Almost as important is the empirical demonstration of the qualitative and quantitative differences between the verbal and exemplar channels of communication. We observed that the exemplar channel is more robust to perceptual confusability of the category structures, i.e., it is more efficient in communicating categories that require higher precision in perceptual decisions. At the same time, the verbal channel is more robust to increases in the dimensionality of the stimuli.

Our simulations show that the proposed *rational category communication* model can capture the main qualitative properties of the empirical data. Additionally, the number of exemplars it chooses to ensure that a student learns a category is in close alignment with empirical data. Most importantly, it captures the difficulties of verbally explaining categories that require high perceptual precision and the robustness of exemplar communication channel to such changes.

Overall, the verbal and exemplar channels of communication have their unique strengths and weaknesses, and their relative efficiency largely depends on the structure of the hypothesis space.

While many of the reported results are preliminary, we hope that both the proposed experimental paradigm and the computational model would facilitate further research into the relative roles of verbal and exemplar information in communicating category structure. To further aid this goal, we make the model implementation openly available.

Lastly, we find that under our experimental settings, the answer to the question of "how many words is a picture worth?" is approximately 7.17.

Acknowledgments

The project was supported by RFBR grant #18-313-00249.

Supplementary materials

Model implementation and other accompanying materials: https://github.com/R-seny/rational-categorization-model

References

Aboody, R., Velez-Ginorio, J., Laurie, R., Santos, L. R., & Jara-Ettinger, J. (2018). When teaching breaks down: Teachers rationally select what information to share, but misrepresent learners' hypothesis spaces. *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*, *1*, 72–77.

Ashby, F. G., Paul, E. J., & Maddox, W. T. (2011). Covis. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (p. 6587). Cambridge University Press. doi: 10.1017/CBO9780511921322.004

Avrahami, J., Kareev, Y., Bogot, Y., Caspi, R., Dunaevsky, S., & Lerner, S. (1997, aug). Teaching by Examples: Implications for the Process of Category Acquisition. *The Quarterly Journal of Experimental Psychology Section A*, *50*(3), 586–606. doi: 10.1080/713755719

DeJong, G., & Mooney, R. (1986). Explanation-based learning: An alternative view. *Machine learning*, *1*(2), 145–176.

Fischer, U. (1994). Learning words from context and dictionaries: An experimental comparison. *Applied Psycholinguistics*, *15*(4), 551574. doi: 10.1017/S0142716400006901

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998. doi: 10.1126/science.1218633

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive science*, *32*(1), 108–154.

Kruschke, J. K. (1992). Alcove: an exemplar-based connectionist model of category learning. *Psychological review*, *99*(1), 22.

Kruschke, J. K. (2008). Models of Categorization. In R. Sun (Ed.), *The cambridge handbook of computational psychology* (pp. 267–301). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511816772.013

Liefooghe, B., Braem, S., & Meiran, N. (2018). The implications and applications of learning via instructions. *Acta Psychologica*, *184*, 1 - 3. (The implications and applications of learning via instructions) doi: https://doi.org/10.1016/j.actpsy.2017.09.015

Lombrozo, T. (2012, mar). Explanation and Abductive Inference. In K. J. Holyoak & R. G. Morrison (Eds.), *The oxford handbook of thinking and reasoning* (chap. 14). Oxford University Press. doi: 10.1093/oxfordhb/9780199734689.013.0014

Longman, C. S., Milton, F., Wills, A. J., & Verbruggen, F. (2018). Transfer of learned category-response associations is modulated by instruction. *Acta psychologica*, *184*, 144–167.

Mitchell, T. M., Keller, R. M., & Kedar-Cabelli, S. T. (1986, Mar 01). Explanation-based generalization: A unifying view. *Machine Learning*, *1*(1), 47–80. doi: 10.1023/A:1022691120807

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading research quarterly*, 233–253.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological review*, *101*(1), 53.

Reed, S. K., & Bolstad, C. A. (1991). Use of Examples and Procedures in Problem Solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(4), 753–766.

Rosedahl, L., & Ashby, F. G. (2018). A new stimulus set for cognitive research.

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive psychology*, *71*, 55–89.

Vinner, S. (2002). The role of definitions in the teaching and learning of mathematics. In *Advanced mathematical thinking* (pp. 65–81). Springer.

Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science*, *34*(5), 776–806.

Williams, J. J., & Lombrozo, T. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive Psychology*, *66*(1), 55–84.

Winston, P. H., Binford, T. O., Katz, B., & Lowry, M. R. (1983). Learning physical descriptions from functional definitions, examples, and precedents. In *Aaai 1983.*