Few-Shot Learning: A Literature Review and Research Proposal

(PSYCH 124J Final Research Paper - June 2025)

Introduction

Humans have the unique ability to generalize rich and complex concepts from very limited input. This ability is known as few-shot learning and it is an important aspect of human cognition that artificial intelligence models have yet to successfully replicate. Understanding how people are able to accomplish few-shot learning is relevant to a variety of tasks like word learning, visual concept learning, classification, and more.

This paper explores three studies that investigate different aspects of few-shot learning in humans. Lake et al. (2019) explores how people learn the meanings of novel pseudowords from limited examples and successfully generalize this knowledge to new sequences. Tiedemann et al. (2022) explores one-shot learning through a drawing task, where participants successfully generate new exemplars of an unfamiliar category after seeing only one example. Malaviya et al. (2022) pushes this field of research further by demonstrating that humans can perform less-than-one-shot learning, where people can successfully categorize stimuli even when the number of examples is less than the number of categories.

Altogether, these studies show the depth and flexibility of human few-shot learning across different modalities. They highlight that humans rely on strong, built-in assumptions that allow us to efficiently and accurately navigate new, sparse learning environments. By comparing these studies, we aim to explore how human learning is applied across different contexts and how this may support development of more human-like machine learning systems in the future.

Research Summary & Analysis

"Human few-shot learning of compositional instructions" (Lake et al., 2019)

Lake et al. (2019) explores the distinctly human ability to learn and use functional concepts from very few examples, also known as few-shot learning. They study this by investigating how people are able to quickly generalize the meaning of completely unfamiliar "words" from limited examples. To minimize reliance on prior linguistic knowledge, the researchers used pseudowords paired with abstract output sequences that couldn't be easily translated into natural phrases in English. This is a careful design condition that allows them to study compositional learning in a controlled environment, without interference from real-world language knowledge.

Participants were given sequence-to-sequence (seq2seq) learning tasks, similar to those used to train machine learning models. Participants were taught instructions through only a few training examples, made up of a sequence of pseudowords matched to a response sequence of colored circles. They were also taught the order in which functions should be applied. They were then asked to generate appropriate circle sequence outputs for new instructions composed of pseudowords in novel sequence patterns. The results showed that participants were able to generalize to new combinations with high accuracy, despite the limited training examples. In contrast, recurrent neural networks (RNNs) trained and tested on the same data and tasks failed to generalize, performing at or near chance when tested on longer novel sequences.

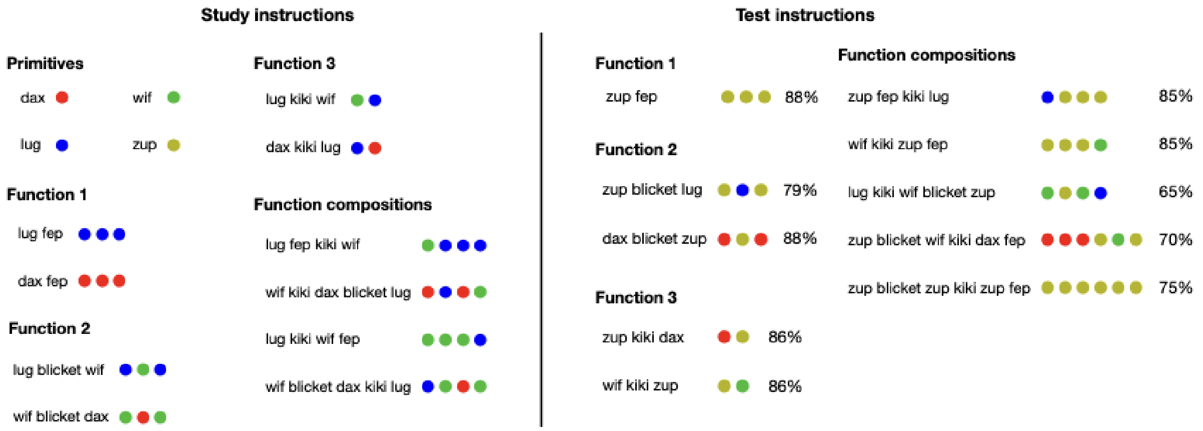


Fig. 1. (Lake et al., 2019) On the left are the instructions provided to participants. The primitives map directly from one pseudoword to one colored circle, while the functions map to transformations. Specifically, Function 1: repeat three times; Function 2: first color, second color, first color; Function 3: second color, first color. On the right are the test instructions. Participants were given the novel pseudoword prompt and were asked to generate the corresponding circle sequence. The percentages are the participants' accuracy on each test instruction, where a sequence was only considered correct if all circles were the right color and in the right order.

An important finding from Experiment 1 was the discovery of three inductive biases that tended to guide participants' generalizations. First, participants tended to assume each pseudoword corresponds to a unique meaning, a phenomenon widely known in linguistics as mutual exclusivity. Second, participants tended to assume that each input word corresponded to exactly one output sequence, without any functional transformations. This preference for one-to-one mappings accounted for 24.4% of all errors in Experiment 1. Third, participants often did iconic concatenation, where they preserved the order of input words in their generated output sequences. This phenomenon is also widely seen in natural language and is an important bias that supports language learning.

In Experiment 2 researchers explored these inductive biases further through a similar experiment design to Experiment 1, but with intentionally ambiguous problems that were compatible with multiple possible generalizations. This allowed the researchers to uncover the participants' default assumptions when the correct answer was not clear. The results confirmed that participants strongly relied on each of the three inductive biases, as participants consistently gave responses in line with the biases. However, a within-subject design was used, meaning the same participants were tested for all three bias conditions. This could lead to earlier trial judgements affecting responses for later experimental trials.

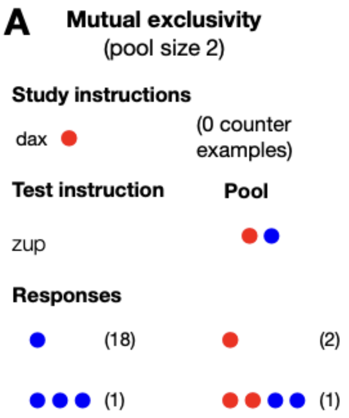


Fig. 2. (Lake et al., 2019) This is an example of one of the trials used to test bias for mutual exclusivity. Participants appear to tend to assume that since “dax” refers to the red circle, “zup” must refer to the blue circle, assuming that two different terms cannot refer to the same thing.

In Experiment 3, participants responded entirely free-form to given pseudoword sequences, without having seen any examples. Even though participants were not given any prior information, most participants defaulted to applying the same inductive biases. Specifically, they assumed unique mappings (mutual exclusivity, one-to-one) for each pseudoword and maintained sequential order (iconic concatenation) when combining them. This result is very compelling because it shows that these biases are present even before participants observe any data, suggesting they may be innate cognitive strategies rather than learned biases. This raises interesting developmental questions about how early these biases can be observed and how they contribute to faster learning through correct generalizations (Singh, 2025). The authors point out that some of these biases, like mutual exclusivity, align with established language learning tendencies observed in children. This connection to developmental literature strengthens the claim that the observed biases are foundational to human learning.

Overall, the study has a strong methodology that is carefully designed to isolate the mechanisms behind human few-shot compositional learning. The results provide strong evidence that humans use structured assumptions to guide their generalizations and enable few-shot learning. These strong inductive biases both support and constrain human generalization and provide a promising direction for future research in human cognition.

“One-shot generalization in humans revealed through a drawing task” (Tiedemann et al., 2022)

This paper by Tiedemann et al. (2022) explores one-shot generalization, the ability to learn a concept from just one example, through a novel drawing task. Instead of merely comparing objects created by the experimenters, participants are prompted to draw their own ‘Variations’ based on a single ‘Exemplar’ shape they are given. This unique design allowed the authors to rigorously examine how individuals generalize from limited information.

The first major finding was that participants were able to synthesize novel objects in a category based on just one exemplar. The exemplars were carefully constructed by the researchers, by combining a central body shape with varying numbers of parts with a wide range of geometric properties. The exemplars were intentionally designed to be diverse and abstract without resembling any real-world items.

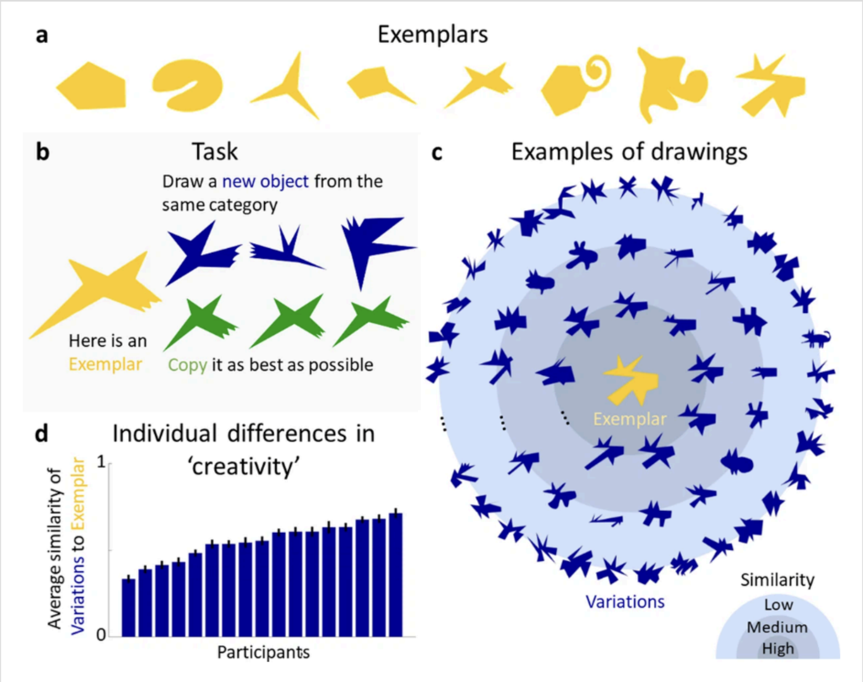

a
Exemplars
b
Task
Draw a new object from the same category
Here is an Exemplar
Copy it as best as possible
c
Examples of drawings
Exemplar
Variations
Similarity
Low
Medium
High
d
Individual differences in 'creativity'
Average similarity of Variations to Exemplar
1
0
Participants

Fig. 3. (Tiedemann et al., 2022) This diagram shows the Exemplars in part a) and examples of the variations participants drew in part c) (in concentric circles by similarity rating).

Another experiment in the study found that a separate group of participants was able to accurately match the drawn variations to the correct exemplar category, with the average percentage of correct classifications being very high (86%). These results suggest that the variations form robust perceptual categories. It also shows that participants were able to identify diagnostic features in exemplars and reproduce them in variations that other observers were able to identify as belonging to the same class.

In their next experiment, the authors further investigated which features were preserved in variations and the degree to which participants agreed about which features were most significant. One feature that consistently carried over across variations was global curvature, which is whether the exemplar had mostly straight or curved lines. Participants tended to preserve this characteristic, producing straight lined variations when the exemplar was straight, and curved variations with the exemplar was curved. Another feature that was preserved was the arrangement and number of parts in the exemplar. While the parts were often altered by size or orientation, their overall structure and organization tended to be retained. This suggests that participants may have segmented the exemplar into parts, modified it, and then mentally reassembled these parts to form variations.

To test this idea, another group of participants was asked to match corresponding parts between exemplars and variations. The results revealed that part correspondence was stronger with the same category than across categories and that part ordering was often preserved. Despite geometric changes, the variations and exemplars shared many of the same parts. Additionally, some parts were perceived as more distinctive than others and were consistently included in the variations. In another experiment, participants showed high agreement about which parts were most distinctive. These parts were most likely to be preserved, indicating that perceived distinctiveness plays a central role in categorization.

In the final experiment, the researchers tested whether distinctive parts directly influenced categorization. They created new stimuli by replacing the most distinctive part of a variation with the most distinctive part of a different exemplar category. This manipulation significantly biased participants' classification judgements towards the category associated with the new distinctive part. These findings suggest that certain parts serve as strong cues for category membership and observers rely on these features to make their categorization.

The authors state that the ability to organize objects into categories at a glance is a fundamental skill that reflects deep perceptual intuitions about how objects vary in the natural world. In everyday life, humans can rapidly generalize from a single example and draw on prior knowledge to predict object properties without needing to relearn them from scratch. The authors believe these one-shot judgements likely depend on internal generative models. In their study, because participants were asked to generate category members themselves, the resulting variations reveal which features they deemed important. People could easily diverge in the features they select, however, the results show a high degree of agreement across participants. This suggests the presence of shared principles in how humans analyze objects and extrapolate new category members from features. Despite the variability in the drawings, most variations were correctly matched to their original categories, supporting the conclusion that participants formed genuine, distinct perceptual categories.

The study's approach of requiring participants to generate their own variations is a compelling way to study the role of internal generative models in human cognition. Participants tended to agree on the most distinctive features, used them in their own variations, and relied on them to classify others' drawings. Some limits to the researchers' approach include constraints from the participants' drawing abilities and possible researcher bias in the hand-designed exemplars. A more rigorous approach could generate the exemplars algorithmically to reduce the chance of unintentionally highlighting certain features (Singh, 2025). The researchers did not fully resolve how the most distinctive features are selected so this could be an area of future research to deepen our understanding of the mechanisms behind one-shot learning.

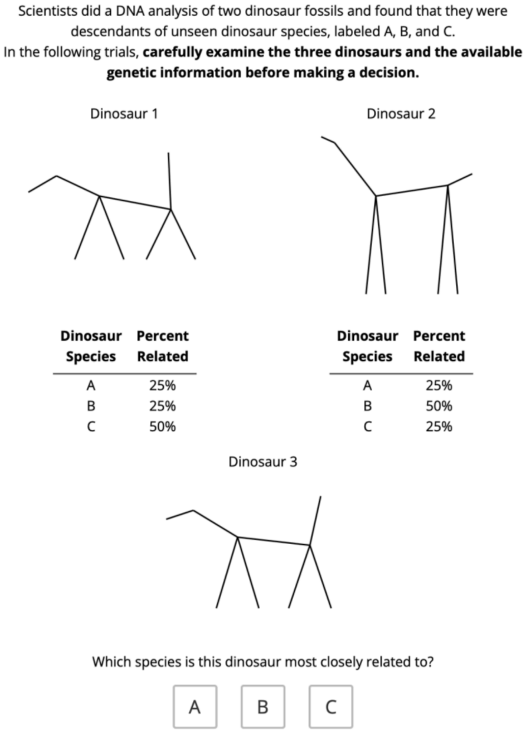
"Can Humans Do Less-Than-One-Shot Learning?" (Malaviya et al., 2022)

Malaviya et al. (2022) investigates exactly how little data humans can learn from. The study focuses on less-than-one-shot learning, in which people successfully learn more categories than the number of examples they have encountered. The researchers introduced a novel experimental paradigm to test this ability in humans. They found that participants were able to learn in extremely data-scarce settings and appeared to rely on prototype-based categorization rather than exemplar-based strategies when making categorical judgements. Additionally, the researchers observed that participants' responses followed systematic patterns, suggesting that humans may use efficient inductive biases that help them with these tasks.

Studying less-than-one-shot learning in humans is particularly challenging because it requires accessing people's internal representations of categories they have never seen before. To address this, the researchers designed an experiment that elicited participants' inferences about unseen categories. The stimuli were stick-figure "dinosaurs", and participants were asked to classify each dinosaur fossil structure into one of three species (A, B, or C). Instead of providing hard category labels, the researchers gave participants "soft labels" in the form of percent genetic similarity to each species. This allowed participants to form inferences based on probabilistic category membership rather than on discrete labels.

$$\begin{bmatrix}0.6\\0.3\\0.1\end{bmatrix} \xrightarrow[\text{and rest to 0}]{\text{Set largest to 1}} \begin{bmatrix}1\\0\\0\end{bmatrix}$$

Fig. 4. (Malaviya et al., 2022) The vector on the left shows soft labels and the vector on the right shows the corresponding hard label. Hard labels provide absolute category membership, soft labels reflect the probability of category membership and may more accurately represent how people make real-life category judgements.



Scientists did a DNA analysis of two dinosaur fossils and found that they were descendants of unseen dinosaur species, labeled A, B, and C.
In the following trials, **carefully examine the three dinosaurs and the available genetic information before making a decision.**

Dinosaur 1

Dinosaur Species	Percent Related
A	25%
B	25%
C	50%

Dinosaur 2

Dinosaur Species	Percent Related
A	25%
B	50%
C	25%

Dinosaur 3

Which species is this dinosaur most closely related to?

Fig. 5. (Malaviya et al., 2022) A screen capture from the experiment, displaying the information given to participants and the prompt.

The two major psychological models commonly used to explain categorization are exemplar models and prototype models. With exemplar models people store all the examples they have seen and classify new stimuli based on their similarity to each of these stored exemplars. By contrast, prototype models create summary representations of categories and classify new items based on their similarity to this prototype instead of comparing it to all instances in the category.

The results of the study showed that participants consistently agree on dinosaur species classifications. Participants were able to accurately infer the underlying feature space that generated the dinosaur stimuli. Based on these findings, the authors argue that participants likely formed prototype-based representations and made classification decisions by comparing new stimuli to these inferred prototypes. The patterns also suggest that participants relied on inductive biases that support efficient learning in situations where data is extremely limited.

It would have been impossible for participants to use an exemplar-based representation to classify the dinosaur skeletons since participants were not given examples of the categories/species to rely on. However, prototype-based categorization allows people to infer the structure of a category and construct a prototype without needing to observe a direct example. A simple example the authors give is, a child might learn that a unicorn is like a horse, but also like a rhino, without ever seeing a unicorn.

More generally, the ability to form detailed conceptual prototypes may work better with the high-information density provided by soft labels, especially in domains where category boundaries are not sharply defined.

The researchers' simple but effective study design closely mirrors how people often learn in natural settings. This method is notable because the use of soft labels is what allows for learning based on feature gradients and enables the learning of categories without any exemplars (less-than-one-shot learning) (Singh, 2025). However, the study’s use of numerical values in the soft labels (genetic similarity values) may weaken its real-world applicability, particularly in the developmental contexts the researchers reference, since the labels require a rudimentary understanding of percentages and genetics (Singh, 2025). A more generalizable design might avoid explicit numbers and/or test children (This participant pool had an avg. age of 36 years old.), who naturally perform few-shot learning without instruction (Singh, 2025).

Comparative Discussion

The three papers, Lake et al. (2019), Tiedemann et al. (2022), and Malaviya et al. (2022) provide a clear testament to the impressive ability of human learning in data scarce environments. These studies explore the cognitive mechanisms that enable humans to generalize from limited information, from one-shot to even less-than-one-shot learning scenarios. Although each paper focuses on a different modality, they all emphasize the human capacity to infer concepts from limited information.

Lake et al. (2019) used pseudowords in a seq-2-seq task to show that people use strong inductive biases to support and constrain their generalizations in understanding linguistic compositional instructions. Tiedemann et al. (2022) showed that people can generalize from a single example to create coherent new category members based on distinctive features of the original. They suggested that participants were not simply copying but actively using internal generative models to guide their responses. Malaviya et al. (2022) pushed this further by using a less-than-one-shot learning classification task which showed that participants were able to infer correct categories even when they had no direct examples from that category. This provided evidence that people can form prototype-based category representations and generalize category membership from indirect relational information.

These studies show that humans can quickly infer concept structure using minimal or even incomplete data. Across the three tasks, people appeared to rely on structured inductive biases (Lake et al., 2019), internal generative models (Tiedemann et al, 2022), and prototype-based categorization (Malaviya et al., 2022) to guide their learning. One way to unify these concepts is to recognize that inductive biases may guide both the internal generative models people use to create new category members and the prototypes they construct to represent categories. In Lake et al. (2019) the researchers conclude there are multiple inductive biases (e.g. mutual exclusivity) that subjects are using to learn the compositional instructions. In Tiedemann et al. (2022) the researchers speculated on what criteria subjects were using to consistently determine the "most distinct" features. This may again be inductive biases, since they could provide foundational assumptions about what kinds of features are important and how objects within a category should relate to each other. These biases could constrain the internal generative models, shaping the way people mentally simulate or draw new examples. Similarly, inductive biases could influence which features are prioritized and preserved in the mental prototypes people create, as seen in Malaviya et al.. Across all three studies, inductive biases could be part of the cognitive processes that support few-shot learning.

Major Conclusions/Issues

- Humans have strong inductive biases that support rapid concept learning.
- Internal generative models enable humans to synthesize novel category members.
- Prototype-based categorization allows for inference beyond seen examples.
- Inductive biases may guide both internal generative models and prototype formation.
- What determines which inductive biases are activated in different tasks? Are they learned during child development or innate? If they are learned, how and when are they learned?

Research Proposal

A key question surrounding this topic is whether or not the inductive biases are innate or flexible strategies that can be shaped through experience. To address this, a future study could investigate whether people can learn inductive biases through targeted training and whether these learned biases can override participants' default, potentially innate biases in concept learning tasks. If inductive biases were fully innate they would likely be resistant to change, even with explicit training otherwise. However, if biases are flexible and shaped by experience, participants should be able to adopt new, experimenter-taught biases and apply them in novel situations. This would suggest that the biases humans use in few-shot learning are learned.

Participants would first complete a baseline categorization task using novel objects that are suspect to multiple biases (e.g. color, shape, symmetry). This would allow researchers to identify a baseline bias for each participant. Then participants would be trained on objects where the categories consistently favor an unusual or non-intuitive bias, while other typical cues are irrelevant. By providing feedback to participants during this phase, participants would be encouraged to adopt the new bias.

After training, participants would be tested on new objects where multiple biases are still valid, but feedback is no longer provided. If inductive biases are learned, participants should now prefer the trained bias, rather than their initial preferences. If the biases are innate or rigid, participants should revert quickly to their baseline biases, ignoring the trained bias. To further test whether the new learned biases influence internal generative models or prototypes, participants could also be asked to generate new objects in the same category (e.g. draw examples).

This research would advance our understanding by clarifying whether inductive biases in categorization and generalization are static constraints or adaptive learning strategies. If biases can be trained it would show that human concept learning is even more flexible than previously known. On the other hand, if participants resist adopting new biases, it would provide further evidence for the innateness of these tendencies and spur more research into the developmental side of inductive biases and learning from limited input. Either way, this research would offer valuable insights into how humans successfully use and potentially learn inductive biases to support few-shot learning.

References

Lake, B., Linzen, T., & Baroni, M. (2019). *Human few-shot learning of compositional instructions*. https://cims.nyu.edu/~brenden/papers/LakeEtAl2019CogSci.pdf

Malaviya, M., Sucholutsky, I., Oktar, K., & Griffiths, T. L. (2022). *Can Humans Do Less-Than-One-Shot Learning?* ArXiv.org. https://arxiv.org/abs/2202.04670

Singh, S. (2025). Annotated Bibliography On Few-Shot Learning. *PSYCH 124J.*

Tiedemann, H., Morgenstern, Y., Schmidt, F., & Fleming, R. W. (2022). One-shot generalization in humans revealed through a drawing task. *ELife*, *11*, e75485. https://doi.org/10.7554/eLife.75485