A unifying computational framework for teaching and active learning

Scott Cheng-Hsin Yang[1], Wai Keen Vong[1], Yue Yu & Patrick Shafto
Department of Mathematics & Computer Science, Rutgers University – Newark

Traditionally learning has been modeled as passively obtaining information or actively exploring the environment. Recent research has introduced models of learning from teachers that involve reasoning about why they have selected particular evidence. We introduce a computational framework that takes a critical step toward unifying active learning and teaching by recognizing that meta reasoning underlying reasoning about others can be applied to reasoning about oneself. The resulting Self-Teaching model captures much of the behavior of information-gain-based active learning with elements of hypothesis-testing-based active learning and can thus be considered as a formalization of active learning within the broader teaching framework. We present simulation experiments that characterize the behavior of the model within three simple and well-investigated learning problems. We conclude by discussing such theory-of-mind-based learning in the context of core cognition and cognitive development.

1 Introduction

We learn by interacting with the world, by choosing what to look at and what to do. Through these actions we reduce uncertainty and accumulate knowledge about the world. We also learn from other people. Sometimes those people are knowledgeable about the world and demonstrate actions with the intent to foster learning, as is the case with informal notions of teaching (Shafto, Goodman, & Griffiths, 2014). Other times, those people may not be fully knowledgeable, but may nevertheless select demonstrations that we observe (Shafto, Eaves, Navarro, & Perfors, 2012). Indeed, people may be no more knowledgeable than ourselves, in which case their actions may come close to uninformed choices of our own (Markant & Gureckis, 2014). To understand learning in realistic contexts, we must understand actions chosen to learn, *active learning*, actions are chosen for others and their implications for learners, *teaching*, as well as both types of choices are integrated in a singular experience of learning.

The literature on active learning and teaching are not well-integrated despite having been the subject of extensive independent and joint research. For instance, there is a robust debate in the educational psychology literature regarding the relative merits of direct instruction and exploratory learning, which are analogs of formal models of teaching and active learning (Bruner, 1961; Kirschner, Sweller, & Clark, 2006; Tobias & Duffy, 2009). Research has focused on identifying which of these is more effective for promoting learning,

[1]Both authors contributed equally to this paper.

with conclusions in favor of exploration (Steffe & Gale, 1995), direct instruction (Kirschner et al., 2006), and rough equivalence (Klahr & Nigam, 2004). Both active learning and teaching have been the subject of research in computational modeling. Active learning is by far the older literature with models dating back to early work by Bruner, Goodnow, and Austin (1956) and sustained recent attention resulting in a variety of computational models which vary in detail but converge around the idea that people choose examples that tend to reduce some globally-defined notion of uncertainty over possible hypotheses (Nelson, 2005; Coenen, Nelson, & Gureckis, 2017; Crupi, Nelson, Meder, Cevolani, & Tentori, 2018). Similarly, recent modeling has formalized teaching and learning from teaching (Shafto et al., 2014; Shafto & Goodman, 2008; Shafto, Goodman, & Frank, 2012; Frank, 2014). These models characterize this form of social learning as a joint inference about the true hypothesis and the process by which the teacher selects data (Shafto et al., 2014; Shafto & Goodman, 2008). Strikingly, although the problems of active learning and teaching are of common interest, no research to our knowledge has investigated whether there exists a common framework in which we may understand both as an expression of a singular learning mechanism.

We explore a unified approach to modeling active learning and teaching. The ultimate goal of both active learning and teaching is for the learner to learn the underlying true hypothesis. The defining feature of active learning is that the learner can intervene the world however he or she likes to discover that hypothesis. However, because the learner does not know the true hypothesis, they do so by choosing the interventions based on an intermediate objective. The literature on computational models of active learning reveals two relatively distinct sets of phenomena summarized by two kinds of objectives. One is that people select interventions with the goal to increase information or minimize uncertainty (Nelson, 2005; Yang, Lengyel, & Wolpert, 2016; Crupi et al., 2018); the other is that people select interventions to test hypothesis with a bias toward confirmation (Wason, 1960; Klayman & Ha, 1987; Navarro & Perfors, 2011; Coenen et al., 2017; Bramley, Dayan, Griffiths, & Lagnado, 2017).

In contrast to active learning, the defining features of teaching is that the learner receives demonstrations of an intervention rather than chooses it and that the demonstration are chosen by the teacher who knows the true hypothesis. Knowing the truth, the teacher's objective is more straight-forwardly to choose demonstrations that will best help the learner infer the true hypothesis after observing the demonstrations. We explore the idea that models of teaching can be transformed to active learning by simply assuming that the correct hypothesis is unknown to the teacher. This has the virtue of preserving the core structural element of theory of mind, while applying it reflexively. In this paper, we mathematically derive this form of *self-teaching* and show that the resulting objective favors interventions that tend to test distinctive hypotheses, i.e., difficult to learn given random observations. We then show through simulation that the selections of the Self-Teaching model and Expected Information Gain (EIG) model are similar, though not always the same. When the two differ significantly, the Self-Teaching model also exhibits behavior of positive-test strategy. Thus, self-teaching not only unifies the phenomena of active learning and teaching; it may also offer insight into how the different phenomena of active learning can be explained.

The goal of the current work is to explore active learning as a form of self-teaching. In Section 2, we introduce the mathematical models of active learning and self-teaching. In

Section 3.1, we describe the simulation of the learning tasks. In Section 3.2, we present the simulation results which characterize the behavior of the Self-Teaching relative to active learning models. In Section 4, we revisit the potential implications of a unified model and describe open questions.

2 Theory

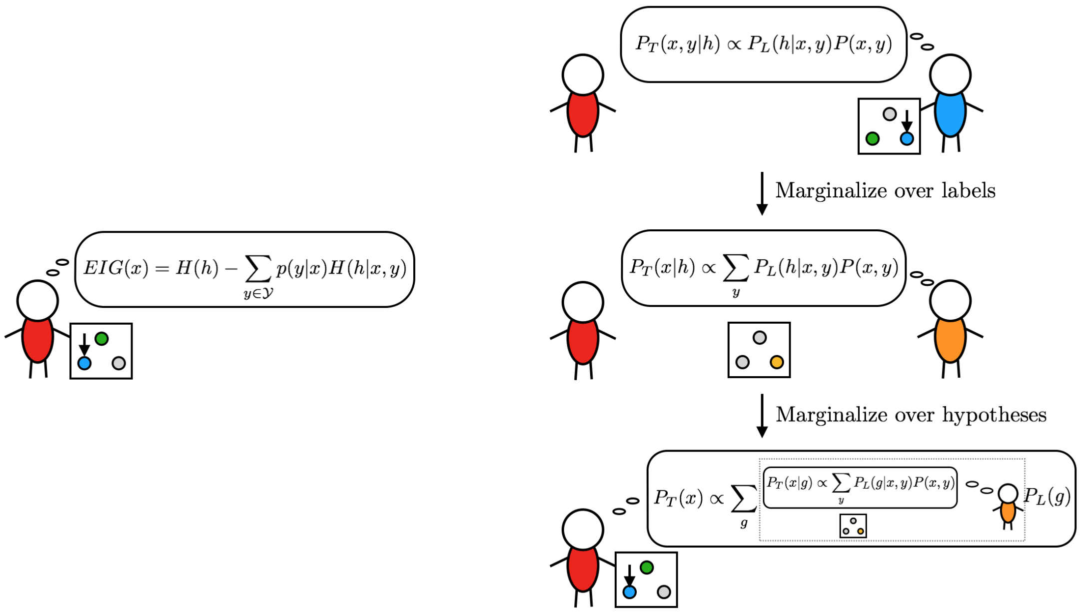
To bridge the gap between these two modes of learning, the focus of our theory is to modify the teaching model of Shafto & Goodman (Shafto & Goodman, 2008; Shafto et al., 2014) in two stages to examine how teaching may be reduced to a form of active learning. In the first stage, we modify the teaching model from providing a demonstration of features and labels to providing instruction on which feature to select (from the top part to the middle part of Figure 1(b)). Mathematically, this involves marginalization over the possible outcomes (labels), which are no longer under the teacher's control. In the second stage, we further modify the teaching model so that the teacher is only as knowledgeable as the learner is. Mathematically, this involves marginalization over the possible hypotheses. Formally, being taught by such a teacher is equivalent to the learner simulating a teacher to teach oneself, and hence the term, self-teaching (from the middle part to the bottom of Figure 1(b)).

Because the learner selects interventions for themselves in self-teaching, self-teaching is a form of active learning. Nevertheless, it is different from information-gain-based active learning in that this single architecture for meta-reasoning underlies learning for oneself through active learning, and learning from others through teaching. From a formal perspective, this reflects a difference in the objective being optimized for. As will be discussed in the following, the Self-Teaching model shares the goal of choosing interventions to help inferring the correct hypothesis with the teaching models, but having no access to the truth, it expresses this goal as a bias to confirm whether the true hypothesis is one that is difficult to learn with random observations. This is in contrast with formalizations of active learning such as the Expected Information Gain model, which aims to reduce expected uncertainty. We highlight both similarities and differences in the predictions of the two models in Section 3.2 after introducing them more formally in this section.

2.1 Notation for inference

We formalize learning as Bayesian inference. The learner's goal is to infer, from amongst all possibilities, the true hypothesis. We denote a hypothesis by h and the space of all hypotheses by $\mathcal{H}$. Each hypothesis is characterized by a particular set of features and labels. We denote a feature and a label by x and y, respectively, and the space of possible features and labels by $\mathcal{X}$ and $\mathcal{Y}$, respectively.

Initially, the learner has a prior over hypotheses $P_L(h)$, where the hypotheses refer to either concepts in the concept learning tasks or graphs in the causal learning task (Section 3.1). For all tasks, the prior is set to be uniform. After observing a feature, label pair (x, y), the learner updates their beliefs in accordance with Bayes' rule: $P_L(h|x, y) \propto P(y|x, h)P_L(h)$. After the learner has observed a data point, the new learner's prior $P_L(h)$ is set to be the current learner's posterior $P_L(h|x, y)$, without loss of generality.



(a) Expected Information Gain (EIG) model (b) From Teaching to Self-Teaching

Figure 1. **Comparison of active learning using either (a) the Expected Information Gain model or (b) the Self-Teaching model (and how it can be derived from teaching)**. In each subfigure, the learner is depicted on the left and teacher (if it exists), is shown on the right. In the EIG model (left), the learner determines which feature x to select by calculating the expected information gain of choosing each feature. On the right, we first illustrate the Teaching-by-Demonstration model (top right), where the teacher demonstrates a feature-label pair (x, y) to the learner. By marginalizing over the possible labels y, this produces the Teaching-by-Instruction model (middle right), where the teacher instructs the learner which feature x to select. Finally, by marginalizing over the hypotheses g, we show how the Self-Teaching model is a form of active learning. Here, the learner simulates a self-teacher (using the Teaching-by-Instruction equations), and marginalizes over all hypotheses to determine which feature to select.

2.2 The Expected Information Gain (EIG) model

In the problem of active learning, a learner must determine how to efficiently sample information from the environment for effective learning. One normative approach to information sampling is to choose which pieces of information would be most helpful at disambiguating between separate hypotheses, known as *information gain* (MacKay, 1992; Nelson, 2005). Using the expected information gain criterion, a learner selects features to intervene upon on the basis of how much information they would gain on average for intervening a particular feature. In order words, the learner selects the feature that maximizes the reduction in uncertainty about the true hypothesis.

In order to calculate the expected information gain, we begin by describing the learner's prior entropy by the Shannon entropy over possible hypotheses:

$$H(h) = \sum_{h \in \mathcal{H}} P_L(h) \log \frac{1}{P_L(h)}. \tag{1}$$

After a learner makes an intervention and observes a feature-label pair (x, y), they can update their beliefs via Bayes' rule to obtain a posterior distribution over hypotheses, $P_L(h|x, y)$. From this, we can calculate the learner's posterior entropy after observing a data point (x, y):

$$H(h|x, y) = \sum_{h \in \mathcal{H}} P_L(h|x, y) \log \frac{1}{P_L(h|x, y)}. \tag{2}$$

We can then calculate the information gain of intervening on a particular feature $x \in \mathcal{X}$ produced a particular label $y \in \mathcal{Y}$, as follows:

$$IG(x, y) = H(h) - H(h|x, y). \tag{3}$$

However, since the set of possible labels that could be observed contains uncertainty, we must calculate the *expected* information gain by weighting the probability of particular labels given the learner selects a particular feature as follows:

$$EIG(x) = H(h) - \sum_{y \in \mathcal{Y}} P_L(y|x)\, H(h|x, y), \tag{4}$$

where $P_L(y|x) = \sum_{h \in \mathcal{H}} P(y|x, h) P_L(h)$ is the predictive distribution over labels. Once the learner has calculated the expected information gain using Equation (4), the learner can choose by either selecting the feature with the highest expected information gain (probability maximizing), or by sampling a feature proportional to its expected information gain (probability matching).

While there exist models of active learning that optimizes other objectives (e.g., Nelson, 2005; Markant, Settles, & Gureckis, 2016), in this paper we focus on EIG. Additionally, we only consider the case where the learner plans one-step into the future, and do not consider the possibility of determining which feature to select by planning multiple steps ahead as some models do (e.g., Bramley, Lagnado, & Speekenbrink, 2015).

2.3 Teaching models

To unify the phenomena of teaching and active learning, we begin with the pedagogical model introduced by Shafto and Goodman (2008) and Shafto et al. (2014), which we call *Teaching-by-Demonstration* here. This model outputs a teaching probability, $P_T(x, y|h)$, which is the probability that the teacher will select a feature-label pair (x, y) to teach an underlying hypothesis h. To move from this Teaching-by-Demonstration model to the Self-Teaching model of active learning, we must introduce uncertainty about the label that will be paired with a choice of feature and the true hypothesis. This is achieved mathematically by marginalizing over possibilities. Each of these marginalizations produces a different kind of teaching model: marginalizing out y gives what we call the Teaching-by-Instruction model with output $P_T(x|h)$ that depends on x and h (Yang & Shafto, 2017); further marginalizing out h gives a *Self-Teaching* model with output $P_T(x)$ that depends on only x as an active learning model should. The objective of the Self-Teaching model is to choose x to confirm whether the true hypothesis is distinctive—unlikely to be identified with random observation—and is different from information-gain-based objectives.

2.3.1 Teaching by Demonstration. In the teaching model, there is a learner and a teacher. We assume that the teacher knows the true underlying hypothesis and gives the learner (x, y) pairs as a demonstration to help the learner infer the true hypothesis. The learner's and teacher's reasoning under such pedagogical setting can be formalized as (Shafto & Goodman, 2008; Shafto et al., 2014):

$$P_L(h|x, y) \propto P_T(x, y|h)\, P_L(h)\,, \tag{5a}$$
$$P_T(x, y|h) \propto P_L(h|x, y)\, P_T(x, y)\,. \tag{5b}$$

The term $P_T(x, y)$ is set to be uniform because the teacher has no reason to favor particular data point *a priori*. To solve this coupled system of equations, one typically begins with computing the learner's posterior with the observation likelihood as usual: $P_{L_0}(h|x, y) \propto P(y|x, h)\, P_L(h)$. This posterior is then fed into Equation (5b) to obtain a $P_{T_0}(x, y|h)$. Then, this $P_{T_0}(x, y|h)$ is fed into Equation (5a) to obtain a new $P_{L_1}(h|x, y)$, which is fed into Equation (5b) to obtain a $P_{T_1}(x, y|h)$. This recursion proceeds until the distributions $P_{L_n}(h|x, y)$ and $P_{T_n}(x, y|h)$ have converged, that is, the distributions do not change with further iteration. What this process describes is that the teacher's objective is to select (x, y) so that the learner would infer h, and that the learner knows that the teacher is selecting demonstration in such a way.

Knowing the true underlying hypothesis, h^*, the teacher selects demonstration (x, y) according to $P_T(x, y|h^*)$, where this is the teacher's selection posterior at convergence. Receiving this demonstration, the learner updates the posterior according to Equation (5a) with this converged teacher's selection probability. The reasoning involves the learner thinking about what the teacher would choose if the true hypothesis were h, for all possible $h \in \mathcal{H}$. Because the learner has reasoned about all possible h and knows that the teacher reasons about things in the same way, when a demonstration of (x, y) is given by the teacher who knows h^*, the learner can identify which hypothesis is h^* effectively by judging which h is likely to lead the teacher to choose that particular demonstration. This recursive and cooperative inference involving a knowledgeable teacher makes the learning the true hypothesis more efficient than active learning (Yang & Shafto, 2017).

2.3.2 Teaching by Instruction. To bring us closer to self-teaching, in which the learner actively selects examples, we consider a Teaching-by-Instruction model. In this model, rather than providing the learner with the datapoint, the teacher instructs the learner about which feature to intervene upon. We introduce this model for purely pedagogical purposes here, although it could be of independent interest in other contexts. Formally, we modify Equation (5) to

$$P_L(h|x,y) \propto P(y|x,h)\, P_T(x|h)\, P_L(h)\,, \tag{6a}$$

$$P_T(x|h) = \sum_{y\in\mathcal{Y}} P_T(x,y|h) \propto \sum_{y\in\mathcal{Y}} P_L(h|x,y)\, P_T(x,y)\,, \tag{6b}$$

where we have decomposed the distribution $P_T(x,y|h)$ to $P(y|x,h)\,P_T(x|h)$ via the chain rule and dropped the subscript T because the label is now provided by the world. The teacher's choice of x now involves marginalization over possible labels in $P_T(x,y|h)$ which is teacher's demonstration selection distribution in Equation (5b). As is in Equation (5), these equations are iterated until convergence, and then the teacher samples x from the converged $P_T(x|h^*)$. Equations (5b) and (6b) differ only in the marginalization of the label, y.

2.3.3 Knowledgeability. So far, we have modeled a teacher with perfect knowledge of the true hypothesis. In order to have a model of self-teaching, we must be able to capture the learner's lack of knowledge. To do this, we formalize *knowledgeability*, which is modeled by the factor $P(g|h)$. In this factor, $h \in \mathcal{H}$ is a hypothetical true hypothesis as is before; the new variable $g \in \mathcal{H}$ is the hypothesis that the teacher is assumed to have in mind given a hypothetical true hypothesis. More concretely, $P(g_j|h_i) = a$ says that if the true hypothesis is h_i, then the teacher thinks the true hypothesis is h_j with probability a. [2]

A teacher with perfect knowledge is described by the *teacher's knowledgeability*, $\delta(g|h)$, where $\delta(g_j|h_i) = 1$ when $j = i$ and is 0 otherwise. This means that when the true hypothesis is h_i, the teacher thinks that h_i is the true hypothesis with probability 1, and this is true for all i. Using this knowledgeability factor, we can write the teacher's selection distribution of Equation (6b) in a more general form:

$$P_T(x|h) = \sum_{g\in\mathcal{H}} P_T(x|g)\, \delta(g|h)\,. \tag{7}$$

This equation says that the marginalization over h with the delta-function simply replaces g with h, which recovers Equation (6b) exactly. This generalization seems trivial here because of the delta function; however, when the knowledgeability factor is less concentrated than the delta function, it offers a way to introduce uncertainty (and/or misconceptions) into the teacher's belief about the true hypothesis.

2.3.4 Self-Teaching. Since there is no actual teacher in self-teaching, the learner does not have access to the true hypothesis. While the teacher's knowledgeability factor is concentrated on the true hypothesis, the *self-teacher's knowledgeability* must reflect learner's current belief. Thus, we set the self-teacher's knowledgeability factor to $P_L(g)$. This says

[2] This can be considered as a generalization of the knowledgeability introduced in Shafto, Eaves, et al. (2012).

that no matter what the (hypothetical) true hypothesis is, the self-teacher believes that the distribution over true hypothesis is the same as the learner's prior, rendering the self-teacher only as knowledgeable as the learner is. Following Equation (7), we add the self-teacher's knowledgeability factor into the teacher's instruction selection probability. This turns Equation (6) into the Self-Teaching equations:

$$P_L(h|x,y) = \frac{P(y|x,h)\, \cancel{P_T(x)}\, P_L(h)}{\sum_{h' \in \mathcal{H}} P(y|x,h')\, \cancel{P_T(x)}\, P_L(h')}, \tag{8a}$$

$$P_T(x) = \sum_{g \in \mathcal{H}} P_T(x|g)\, P_L(g) = \sum_{g \in \mathcal{H}} \sum_{y \in \mathcal{Y}} P_L(g|x,y)\, P_T(x,y)\, \frac{P_L(g)}{Z(g)}. \tag{8b}$$

In Equation (8b), the first equality has the same form as Equation (7) but includes the self-teacher's knowledgeability. To obtain the second equality, we expand the term $P_T(x|g)$ according to Equation (6b). The term, $Z(g) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} P_L(g|x,y)\, P_T(x,y)$, is the normalizing constant that comes from changing the proportionality in Equation (6b) to an equality in Equation (8b).

Note that the self-teacher's knowledgeability, $P_L(g)$, does not depend on h. Thus, after the marginalization of g, the self-teacher's selection of Equation (8b) depends only on x and not on h. Hence, it is this marginalization with the self-teacher's knowledgeability factor that turned the $P_T(x|h)$ in the Teaching-by-Instruction model into $P_T(x)$. As in the teaching framework, we can compute the teacher's selection probability in Equation (8b) by feeding the result into the learner's inference in Equation (8a). However, because the teacher's selection does not depend on h, this factor cancels out in Equation (8a). This indicates that the learner gains no extra information about h from reasoning about the self-teacher's selection, which is what one would expect. This is also reflected by the fact that the self-teacher samples features to intervene according to Equation (8b), which does not depend on h and hence cannot depend on the true underlying hypothesis, h^*.

As for other models of active learning, the Self-Teaching model predicts the choices one would make to learn. However, the objective of Self-Teaching is different from the EIG model. To explore the behavior of Equation (8b), recall that $P_T(x,y)$ is uniform. This means that if the ratio $P_L(g)\,/Z(g)$ is uniform over g, then the self-teacher's selection distribution, $P_T(x)$, will be uniform over x.[3] The term $Z(g)$ is the average probability that the learner will infer g as the true hypothesis if all pairs of (x,y) have equal probability to be observed. Thus, $Z(g)$ can be thought of as how "common" g is. In this view, the Self-Teaching model is a hypothesis-testing model with the objective of confirming whether the underlying true hypothesis has high *expected relative distinctiveness*, that is, it selects x so that on average the learner will more likely observe distinctive (as measure by the $Z(g)^{-1}$) hypotheses relative to the prior.

2.4 Remarks on the models

The learner's inference equations in the teaching models—Equations (5a), (6a), and (8a)—show that different learning settings make use of different sources of knowledge. In

[3]When this ratio is uniform, it can be taken outside the sums. The same is true of $P_T(x,y)$. Then, the sum $\sum_{g \in \mathcal{H}} P_L(g|x,y) = 1$ for all x and y by definition, resulting in all the factors being uniform.

the Teaching-by-Demonstration model, in addition to the learner's prior, Equation (5a) involves only $P_T(x, y|h)$, suggesting that the teacher is the only source of information. In the Teaching-by-Instruction model, Equation (6a) involves $P_T(x|h)$ and $P(y|x, h)$, suggesting that the information comes partly from the teacher and partly from interacting with the world. In self-teaching, Equation (6a) involves only $P(y|x, h)$, indicating that the world is the only source of information. This is true for the EIG and other active learning models, making the two equivalent in terms of source of information for inference.

A similar observation can be made for the feature selection equations. The selections of intervention according to Equations (5b) and (6b) in the Teaching-by-Demonstration and -by-Instruction models both involve the underlying true hypothesis h^*, whereas the selections according to Equation (4) and Equation (8b) do not. Thus, in a teaching setting, the learner can benefit from the teacher's knowledge of the true hypothesis.

While the Self-Teaching model and EIG model are similar in the above mentioned ways, their selection objectives show at least two important differences. First, the EIG model, as well as other conventional active learning models, only reason about the world, whereas the Self-Teaching model involves a meta-cognitive component reasoning about one's own reasoning. Mathematically, this is expressed by the use of $P_T(x|g)$ and $P_L(g)$ in Equation (8b), indicating that the learner is simulating themself (the $P_L(g)$ term) as a teacher (the $P_T(x|g)$ term) to reason about themself as a learner. In contrast, the EIG model is solely a function of the learner's prior and observation likelihood. Second, the reasoning behind the two approaches is quite different. The Self-Teaching model inherits reasoning about individual hypotheses from the teaching model (see the end of Section 2.3.1). This resembles a kind of hypothesis testing (see the end of Section 2.3.4), where features are chosen based on their expected relative distinctiveness with respect to individual hypotheses. In contrast, the EIG model is based on entropy, which is a quantity derived from averaging over hypotheses. Thus, the Self-Teaching model reasons more on the level of individual hypothesis, whereas the information-gain model reasons more on the level of global changes in the beliefs about hypotheses. Interestingly, despite this difference in reasoning and differences in mathematical expression, [4] their intervention selection behaviors are sometimes quite similar. In the next Section, we characterize the similarities and differences through simulation.

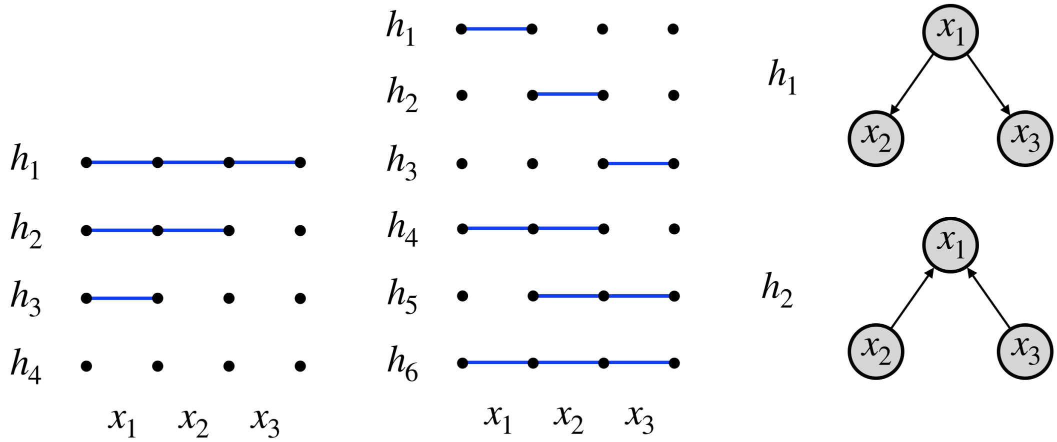
3 Simulations

3.1 Tasks

We illustrate the Self-Teaching model with three tasks: two variants of a concept learning task and a causal learning task. In each of these tasks, we compare and contrast the predictions of the EIG model along with our Self-Teaching model.

3.1.1 Concept Learning. The first two tasks we consider are concept learning tasks, where the set of possible concepts are various line segments on an interval as illustrated in Figure 2(a) and Figure 2(b). The goal of the learner is to discover the true hypothesis by querying features from different positions along the line to determine whether

[4] The Self-Teaching model involves only Bayes' rule and the sum and product rules of probability, whereas EIG requires more math operations, namely, log and minus.



(a) Boundary task (b) Line task (c) Causal learning task

Figure 2. **Examples of the three tasks to compare the expected-information-gain and Self-Teaching models**. We explore two kinds of concept learning which we call the *boundary task* and the *line task*, which both involve concepts of varying line segments on a one-dimensional line. In both concept learning tasks, the set of features x_i refer to the different positions along the line, while the labels y_i indicate whether or not the line segment exists for a particular hypothesis. In the boundary game, the set of hypotheses is restricted to lines where there is a single boundary to disambiguate between hypotheses. On the other hand, the line task includes all possible single line concepts for a given number of features. In the *causal learning task*, the hypotheses h are two different causal graphs. The set of features x_i in this task refer to which node the learner chooses to intervene on, which produces a label y_i corresponding to the states of all three nodes in the observed graph.

that point is inside or outside of the true hypothesis. We consider two variants of the concept learning task which we call the *boundary task* and the *line task*.

In the boundary task, the hypothesis space consists of all of the unique concepts that are defined by the location of their boundary (e.g. Markant & Gureckis, 2014). Intuitively, a unique hypothesis has been identified when the learner observes two neighboring positions with different labels. In the line task, the hypothesis space consists of all possible lines with n features (Tenenbaum, 1999; Shafto & Goodman, 2008). While the set of concepts are similar in these two tasks, the value of selecting different features differs in either task due to the presence of the other hypotheses under consideration. For both concept learning tasks, the learner can ask for only one label at a time.

We use the following notation for describing the concept learning tasks. First, each hypothesis h in the concept learning task consists of a set of features $\mathcal{X} = \{x_i | x_i = i, i = 1, \ldots, n\}$ and labels $\{y_i | y_i \in \{0, 1\}, i \in \{1, \ldots, n\}\}$, where each feature, x_i, refers to a position along the line segment, and each label, y_i, refers to whether the line segment for a given hypothesis exists at that position or not. Thus, each hypothesis can be described as a set of data points $h = \{(x_i, y_i)\}_{i=1}^{n}$. For this tasks, the likelihood $P(y|x, h)$ is given by weak sampling, and is equal to 1 when $x = x_i, y = y_i$ and 0 when $x = x_i, y \neq y_i$ for the given h.

3.1.2 Causal Learning. The third task we explore is a causal learning problem. In this task, the learner is presented with a causal system where the connections between the nodes are not shown, and the learner must determine which of two possible causal graphs is the true causal graph. Each causal graph consists of three nodes and the nodes are connected in different configurations (see Figure 2(c) for one example). The learner can choose to intervene on one of the three nodes, which will turn that particular node on, and will also cause other nodes that are connected downstream of the intervened node to turn on with high probability.[5] Depending on the true underlying graph, intervening on different nodes will produce different patterns of nodes turning on, which can help the learner determine which of the two graphs is the true hypothesis. We consider 27 possible pairs of causal graphs, matching the set used in Coenen, Rehder, and Gureckis (2015), to compare the predictions of the EIG and Self-Teaching models for this task.

The notation for the causal learning tasks is similar to the concept learning tasks, albeit with a few minor differences. The hypotheses h in the causal learning tasks refer to the different graphs, and each pair of graphs form a hypothesis space $\mathcal{H}$. The features in this task, x, refer to the three nodes the learner can intervene on, thus $x \in \{x_1, x_2, x_3\}$. The corresponding label, y, refer to the state of the causal graph after intervening on feature x_i. Thus, if the learner intervened on x_1, and this caused both x_2 and x_3 to turn on, then the corresponding label would be $y = \{\text{intervenedOn, observedOn, observedOn}\}$. The set of possible labels for the causal learning task include the four possible sets of observations on the two other nodes when intervening on the third, leading to twelve possible values of y. For this task, the likelihood $P(y|x, h)$ is the likelihood of observing a particular combination of states for all three nodes, as given by the joint probability over the directed graphical model.

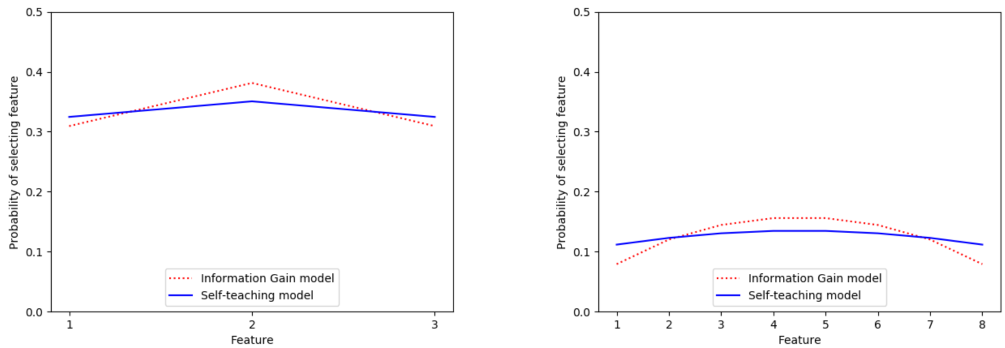
3.2 Results

In this section, we explore via simulation how the predictions of our proposed Self-Teaching model fares when compared to EIG model. In particular, we explore similarities and differences between these models in their choice of particular features to intervene upon.[6]

For the boundary task, we ran simulations using both three features and eight features as shown in Figure 3. The results show that the EIG model and Self-Teaching both mildly favour the selection of features towards the center, rather than the ends. This is true both in the three feature case (Figure 3(a)) where the best choice according to either model is the central feature, and the eight feature case (Figure 3(b)), where both models show a preference for the two middle features. Another observation in both of these figures is that the EIG model shows a slightly stronger preference for selecting what it thinks are the best

[5]For our simulations, the probability that an active parent node would turn on its direct descendants was set to be 0.8. Additionally, while some versions of this task also assume some small chance of a node spontaneously turning on by itself, we did not allow for this possibility in our simulations.

[6]The source code to run the simulations and produce the Figures in the Results section can be found at `https://github.com/CoDaS-Lab/self-teaching`.



(a) Three features (b) Eight features

Figure 3. **Probability of selecting each feature for the boundary task with three and eight features**. The horizontal axes represent the feature position along the line for each concept. The vertical axis represents the probability of selecting each feature. The predictions of the Self-Teaching model are shown in blue, while the EIG model is in red. Both the EIG and Self-Teaching models favour the selection of features in the middle for the boundary task for both three and eight features, with the EIG model showing a stronger preference.

features, in comparison to the Self-Teaching model. We will see more examples of this for the other simulations.

To further explore the behaviour of both models in the boundary task, we conducted an additional set of simulations considering how the EIG model and Self-Teaching model would determine which feature to select after observing a single data point. The plots showing the four different possibilities are shown in Figure 4.[7] For the four sets of observations, in each case the results show that the EIG model and Self-Teaching both favour the same sets of features, with the EIG model again showing a stronger preference.

To further support the conclusion that the Self-Teaching model accords with the behaviour of the EIG model, we consider predictions on a related, but different concept learning task. Results for the line task are shown in Figure 5, where we considered hypothesis spaces that again consisted of either three or eight features. The results from these two simulations show different, yet interesting patterns of predictions that are qualitatively different from the boundary task. First, in the three feature case as shown in Figure 5(a), both the Self-Teaching and EIG models show a preference for the features on the ends, rather than in the center like the boundary task results in Figure 3. This suggests that Self-Teaching and EIG exhibit a similar sensitivity to the set of hypotheses and assign similar values to the features. Indeed, examination of the probabilities of selecting each feature in a line task with eight features in Figure 5(b) reveals interesting differences compared to the boundary task with eight features in Figure 3(b). Here, we find that the EIG model peaks

[7] We do not show the two sets of data points which end up discovering the true hypothesis from a single observation. After the learner observes $x = 1, y = 1$, or $x = 3, y = 0$, the only consistent hypothesis remaining is h_1 or h_4 respectively as depicted in Figure 2(a), and thus there is no additional benefit to exploring the predictions of either model after having observed either of these two cases.

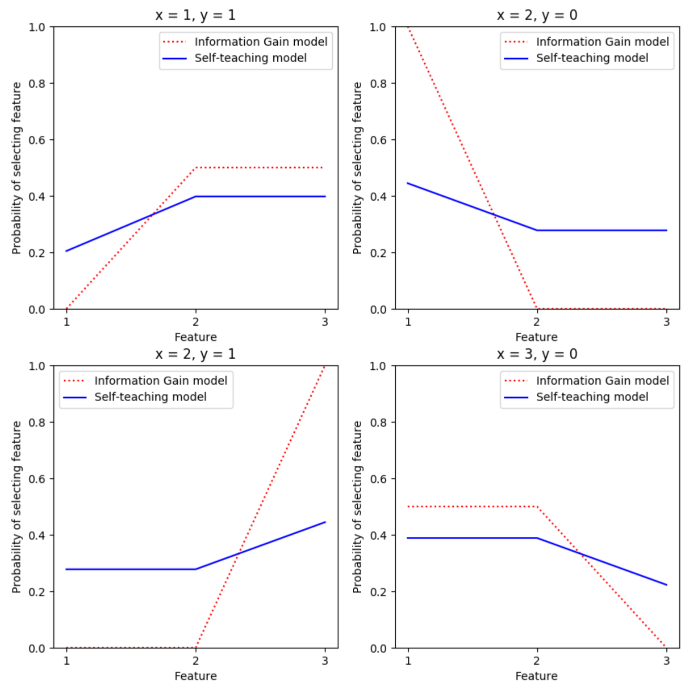
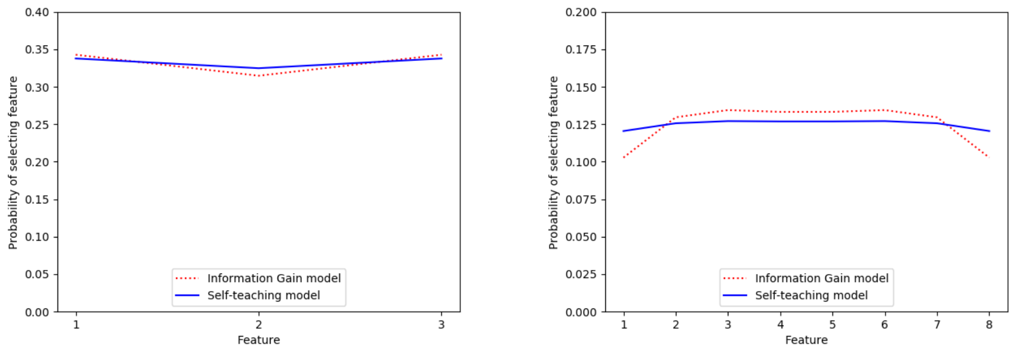


Figure 4. **Probability of selecting each feature for the boundary task with three features after observing a single data point**. The title of each figure is the first data point observed by the model. Two conditions are omitted, as observing either of these cases would lead the learner to infer the true hypothesis after a single step. After a single observation, both the Self-Teaching model (blue) and the EIG model (red) show the same qualitative behaviour in determining which feature to select next, with the EIG model showing stronger preferences for particular features.

in two points away from the center, suggesting selecting the features at those points would provide more information than the center. Similarly, we also see that Self-Teaching predicts the same qualitative behaviour, although the differences between selecting features is less pronounced. The EIG model show a greater preference towards the best features compared to Self-Teaching because the subtraction operation allows EIG to be zero for some features. The Self-Teaching model would never assign zero probability to any feature, resulting in more attenuated preferences (Figure 4).

So far, our results show that the qualitative behaviour of Self-Teaching matches the EIG model in two different tasks involving line concepts. However, is this always the case, or are there situations where the predictions of the Self-Teaching model differ from the EIG model?

To explore this possibility, we also compared the EIG and Self-Teaching models in a causal learning setting. In particular, we examined the predictions of both models across the



(a) Three features (b) Eight features

Figure 5. **Probability of selecting each feature in the line task with varying numbers of features features.** The horizontal axes represent different features for each concept. The vertical axis represents the probability of selecting the feature. Here, the qualitative pattern of predictions differs from the boundary game for both the three feature case and the eight feature case. In the three feature case, both the Self-Teaching model (blue) and the EIG model (red) have a preference for choosing the features on the ends, rather than the central feature like the boundary task. Additionally, for the eight feature case, instead of a single peak at the center, both the Self-Teaching model (blue) and the EIG model (red) have a preference at two particular peaks away from the center.

27 problems used in Coenen et al. (2015), as they provided a challenging test-bed of active learning problems within the domain of causal learning. In particular, neither of the two models (EIG and Positive-Test Strategy) considered in the paper alone explains the human data particularly well, and we were particularly interested in whether the Self-Teaching model could unify the various phenomena in this domain.

Participants were presented with an unknown causal system (by showing three nodes, but not the particular edges connecting the three nodes), and were asked to figure out how the causal system worked. Participants were presented with two possible hypotheses to explain the causal system (by showing two different sets of connecting edges between the nodes). They were asked to determine which of the two possible hypotheses was the true underlying causal system, and could do so by choosing one of the three nodes to intervene on and observe how the system operated. For each problem, the proportion that each node was selected across all participants was noted by Coenen et al. (2015).

The behavioural data collected from Coenen et al. (2015) was then compared to two different active learning models: the EIG model which we have described earlier, and a Positive-Test Strategy model (PTS), which is reminiscent of confirmation bias strategies for active learning. In this particular task, the behaviour of the PTS model favours nodes which turn on as many other nodes as possible, by calculating the following:

$$\mathrm{PTS}_x = \max_x \Big[\frac{\mathrm{DescendantLinks}_{x,h}}{\mathrm{TotalLinks}_h}\Big]. \tag{9}$$

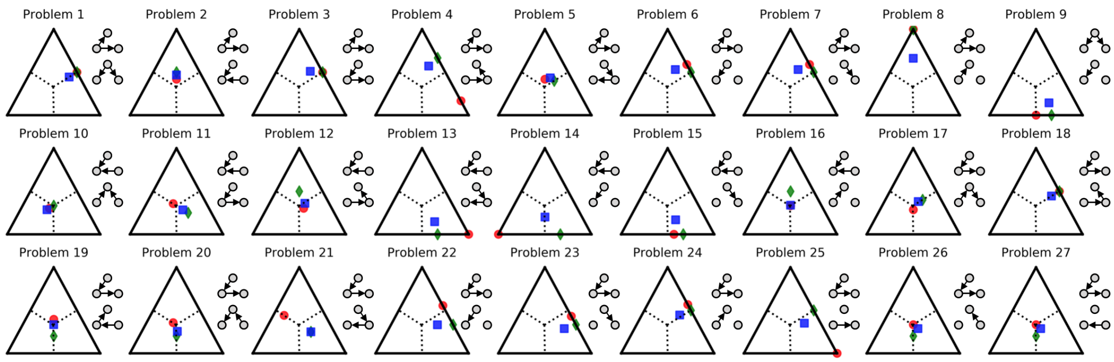


Figure 6. **The predictions of the Self-Teaching model (blue square), EIG model (red circle) and Positive-Test Strategy model (green diamond) for the different causal learning problems.** Each of the diagrams represents a probability simplex over the three nodes to intervene upon, along with the two hypotheses to be considered in the top right of each diagram. A point in the center of the simplex indicates indifference towards selecting any of the three nodes, while points that are closer to each corner represent a stronger preference for selecting one node over the others. The dotted lines indicate the boundaries for favouring one feature over the other two, partitioning the simplex into three distinct sectors. Our results show that for most of the 27 problems, the predictions of the Self-Teaching model for which node to intervene matches the EIG model (see text for more details). However, in a number of cases, the Self-Teaching model prefers intervening on a different node. We compared this to another model of active learning (Positive-Test Strategy) and found that this made similar predictions in these non-matching cases, suggesting that Self-Teaching mimics the behaviour of both models under different circumstances.

Overall, the results from Coenen et al. (2015) showed that while both models captured participant behaviour well in some of the problems, neither of the two active learning models alone were able to capture participant's responses across all of the of problems.

We computed the predictions for the Self-Teaching model (along with the EIG and PTS models) across the set of 27 causal learning problems, and results for these simulations are shown in Figure 6, with the two causal graphs under consideration for each problem shown in the top right of each plot and the predictions from each of the three models plotted in different colours and symbols. We say that the behaviour of the Self-Teaching model differs from the EIG model if the predictions diverge in a noticeable manner, and there are two cases where this occurs. First, is when the predictions of the EIG model favours selecting a single feature with high probability such that the point lies on one of the vertices (Problems 8, 13 and 25). In these instances, the Self-Teaching model favours selecting the same feature as predicted by the EIG model (as both points lie within the same sector), but not as strongly. The second case is when Self-Teaching favours selecting a different feature than the EIG model as evidenced by their predictions lying in two different sectors (Problems 4, 14, 21 and 22). In the latter cases where Self-Teaching prefers a different feature than the EIG model, the predictions of the PTS model also prefer the same feature that Self-Teaching does (except for Problem 14). Apart from these few cases, we consider the predictions of the EIG model and the Self-Teaching model to be matching, and find that there is a strong amount of similarity between these two active learning models.

4 Discussion

People learn from active exploration and from other people. Yet, research on active learning and teaching has largely been conducted independently or contrastively, and to our knowledge there exist no unified models of these basic cognitive phenomena. These models differ in terms of both the learning objective and learning architecture, making unification an interesting and challenging problem. We have presented an integration through formalizing active learning as self-teaching. In our approach, active learning is simply teaching when one cannot predict the outcome of one's actions and does not know the true hypothesis. Mathematically, this corresponds to adding two simple marginalization steps to a previously proposed teaching model, with the resulting model being a formalization of active learning as self-teaching.

We compare Self-Teaching to a representative model from the active learning literature, Expected Information Gain (EIG). Mathematically, the models yield quite different selection functions (see Sections 2.2, 2.3.4, and 2.4). Indeed, the differences are substantial enough that direct mathematical comparison is non-trivial. Inspection of the Self-Teaching model, however, shows that it seeks data that, in expectation, yield changes in beliefs about specific, distinctive hypotheses relative to random sampling. This expected relative distinctivenss differs from expected information gain in focusing on 1) specific hypotheses rather than the entire distribution and 2) hypotheses that are relatively distinctive in the sense that one would be unlikely to learn about through by random sampling rather than the divergence between prior to posterior.

A well-known challenge for existing models of active learning is that there are two relatively distinct sets of phenomena that have been observed empirically. There exist robust literatures in which people's selection of interventions is approximated by models that aim to increase information or minimize uncertainty (Nelson, 2005; Crupi et al., 2018) and where people's selection of interventions has been characterized as models of hypothesis testing with a bias toward confirmation (Wason, 1960; Klayman & Ha, 1987; Navarro & Perfors, 2011). The EIG model can explain the former but not the latter (Markant & Gureckis, 2014; Coenen et al., 2015). Interestingly, the Self-Teaching model qualitatively captures both phenomena through expected relative distinctiveness. Our results show close correspondence with the EIG model on some tasks (Figures 3–5) and closer to a mixture of information gain and positive-test strategies in others (Figure 6). Interestingly, this intermediate approach qualitatively agrees with the conclusions of recently published research (Coenen et al., 2015). This suggests that Self-Teaching can capture different forms of active learning that previously required different mathematical architecture to formalize.

Unifying active learning and teaching as meta-reasoning has potentially far-reaching implications. A model that unifies the two would allow interpolation between these two extremes by manipulating the teacher's knowledgeability. These interpolations allow direct comparisons between teaching from teachers of different knowledge levels, active learning, guided learning (Weisberg, Hirsh-Pasek, & Golinkoff, 2013), and collaborative learning between peers (Dillenbourg, 1999). More generally, unifying models of teaching and active learning may provide a first step towards unifying models of cognition about others (i.e., social cognition; (Shafto, Eaves, et al., 2012; Eaves & Shafto, 2012; Baker, Saxe, & Tenenbaum, 2011) and cognition about oneself (i.e., meta-cognition) in a single meta-reasoning architeucture. Research in psychology and neural science has long speculated about a close link between social cognition and meta-cognition including the possibility of common developmental trajectory and neural substrates (Meltzoff, 2005). Work remains to explore how the current framework relates to more varieties of active learning models in the literature (e.g. Austerweil & Griffiths, 2011; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003) and how much of human behavior it captures. Overall, based on the simulation results, there is reason to be optimistic that the different models and model behaviors can be tied together to yield an integrated account of human learning from people and the world.

Acknowledgements

This material is based on research sponsored by the Air Force Research Laboratory and DARPA under agreement number FA8750-17-2-0146 to P.S. and S.Y. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. This research was also supported in part by NSF SMA-1640816 to P.S.

References

Austerweil, J. L., & Griffiths, T. L. (2011). A rational model of the effects of distributional information on feature learning. *Cognitive Psychology*, *63*(4), 173–209.

Baker, C., Saxe, R., & Tenenbaum, J. (2011). Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society.*

Bramley, N. R., Dayan, P., Griffiths, T. L., & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, *124*(3), 301–338.

Bramley, N. R., Lagnado, D. A., & Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through sequences of interventions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 708–731.

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, *31*, 21–32.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). A study of thinking. *New York*.

Coenen, A., Nelson, J. D., & Gureckis, T. (2017). Asking the right questions about human inquiry. *PsyArXiv*. doi: 10.31234/osf.io/h457v

Coenen, A., Rehder, B., & Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive Psychology*, *79*, 102–133.

Crupi, V., Nelson, J. D., Meder, B., Cevolani, G., & Tentori, K. (2018). Generalized information theory meets human cognition: Introducing a unified framework to model uncertainty and information search. *Cognitive Science*, 1–47.

Dillenbourg, P. (1999). *Collaborative learning: Cognitive and computational approaches.* ERIC.

Eaves, B., & Shafto, P. (2012). Unifying pedagogical reasoning and epistemic trust. *Advances in Child Development and Behavior*, *43*, 295–319.

Frank, M. (2014). Modeling the dynamics of classroom education using teaching games. In *Proceedings of 36th Annual Conference of the Cognitive Science Society.*

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, *15*(10), 661–667.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211–228.
MacKay, D. J. (1992). Information-based objective functions for active data selection. *Neural Computation*, *4*(4), 590–604.
Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology*, *143*(1), 94–122.
Markant, D. B., Settles, B., & Gureckis, T. M. (2016). Self-directed learning favors local, rather than global, uncertainty. *Cognitive Science*, *40*(1), 100–120.
Meltzoff, A. N. (2005). Imitation and other minds: The "like me" hypothesis. *Perspectives on Imitation: From Neuroscience to Social Science*, *2*, 55–77.
Navarro, D. J., & Perfors, A. F. (2011). Hypothesis generation, sparse categories, and the positive test strategy. *Psychological Review*, *118*(1), 120–134.
Nelson, J. D. (2005). Finding useful questions: on bayesian diagnosticity, probability, impact, and information gain. *Psychological Review*, *112*(4), 979–999.
Shafto, P., Eaves, B., Navarro, D. J., & Perfors, A. (2012). Epistemic trust: Modeling children's reasoning about others' knowledge and intent. *Developmental Science*, *15*(3), 436–447.
Shafto, P., & Goodman, N. (2008). Teaching games: Statistical sampling assumptions for learning in pedagogical situations. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1632–1637).
Shafto, P., Goodman, N. D., & Frank, M. C. (2012). Learning from others: The consequences of psychological reasoning for human learning. *Perspectives on Psychological Science*, *7*(4), 341–351.
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.
Steffe, L. P., & Gale, J. E. (1995). *Constructivism in Education.* Lawrence Erlbaum Hillsdale.
Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, *27*(3), 453–489.
Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In *Advances in Neural Information Processing Systems* (pp. 59–65). MIT Press.
Tobias, S., & Duffy, T. M. (2009). *Constructivist instruction: Success or failure?* Routledge.
Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*(3), 129–140.
Weisberg, D. S., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Guided play: Where curricular goals meet a playful pedagogy. *Mind, Brain, and Education*, *7*(2), 104–112.
Yang, S. C.-H., Lengyel, M., & Wolpert, D. M. (2016). Active sensing in the categorization of visual patterns. *Elife*, *5*, 1–22.
Yang, S. C.-H., & Shafto, P. (2017). Teaching versus active learning: A computational analysis of conditions that affect learning. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

Appendix

A worked example

Below, we show the calculations for both the self-teaching model and information gain model for the probability of selecting each feature for the boundary task with three features as shown in Figure 2(a). In this task, there are four hypotheses and the **learner's prior** is set to be uniform for each hypothesis: $P(h) = 1/4$. The **observation likelihood**, $P(y|x, h)$, is shown in Table A1(a).

Both the expected-information-gain and self-teaching models require calculating the **learner's posterior** probabilities, which is given by $P_L(h|x, y) = \frac{P_L(y|x,h)P_L(h)}{\sum_{h' \in \mathcal{H}} P_L(y|x,h')P_L(h')}$. As

Table A1
***(a)**. The observation likelihood, $P(y|x,h)$. The last column of this table computes $\sum_{h'\in\mathcal{H}} P_L(y|x,h')$, which is useful for computing the posterior. **(b)**. The learner's posterior probabilities, $P_L(h|x,y)$. Since the learner's prior is constant in h, this can be obtained by taking each element of the likelihood table and divide that by the sum of the corresponding row recorded in the last column. The last row of this table is the sum of each column, which will be useful for self-teaching.*

$P(y\|x,h)$	h_1	h_2	h_3	h_4	$\sum_h$
$x=1,y=0$	0	1	1	1	3
$x=1,y=1$	1	0	0	0	1
$x=2,y=0$	0	0	1	1	2
$x=2,y=1$	1	1	0	0	2
$x=3,y=0$	0	0	0	1	1
$x=3,y=1$	1	1	1	0	3

(a) *The observation likelihood*

$P_L(h\|x,y)$	h_1	h_2	h_3	h_4
$x=1,y=0$	0	$1/3$	$1/3$	$1/3$
$x=1,y=1$	1	0	0	0
$x=2,y=0$	0	0	$1/2$	$1/2$
$x=2,y=1$	$1/2$	$1/2$	0	0
$x=3,y=0$	0	0	0	1
$x=3,y=1$	$1/3$	$1/3$	$1/3$	0
$\sum_x \sum_y$	$11/6$	$7/6$	$7/6$	$11/6$

(b) *The learner's posterior*

a concrete example, consider the learner's posterior for h_2, after observing that $x=1$ and $y=0$:

$$P_L(h_2|x=1,y=0) = \frac{P_L(y=0|x=1,h_2)\,P_L(h_2)}{\sum_{h'\in\mathcal{H}} P_L(y=0|x=1,h')\,P_L(h')} = \frac{1\cdot\frac{1}{3}}{0\cdot\frac{1}{3}+1\cdot\frac{1}{3}+1\cdot\frac{1}{3}+1\cdot\frac{1}{3}} = \frac{1}{3}.$$

The learner's posterior probabilities for each hypothesis for every possible observation is recorded in Table A1(b).

Self-teaching

To compute the self-teaching selection probability, we first compute the **teacher's demonstration selection probability** according to Equation (5b): $P_T(x,y|h) = \frac{P_L(h|x,y)P_T(x,y)}{\sum_{x'\in\mathcal{X}}\sum_{y'\in\mathcal{Y}} P_L(h|x',y')P_T(x',y')}$. Since there are 6 combinations of (x,y), we set the **teacher's selection prior** to be uniform, $P_T(x,y) = 1/6$ for all x,y pairs. Note that this $P_T(x,y)$ always cancels out, but we still show it for clarity. Now, we can use the posterior probabilities in Table A1(b) to compute the teacher's demonstration selection probability. Continuing our example, suppose the true hypothesis is h_2, and the teacher wants to demonstrate the data point $x=1,y=0$:

$$\begin{aligned} P_T(x=1,y=0|h_2) &= \frac{P_L(h_2|x=1,y=0)\,P_T(x,y)}{\sum_{x'\in\mathcal{X}}\sum_{y'\in\mathcal{Y}} P_L(h_2|x',y')\,P_T(x',y')} \\ &= \frac{\frac{1}{3}\cdot\frac{1}{6}}{\frac{1}{3}\cdot\frac{1}{6}+0\cdot\frac{1}{6}+0\cdot\frac{1}{6}+\frac{1}{2}\cdot\frac{1}{6}+0\cdot\frac{1}{6}+\frac{1}{3}\cdot\frac{1}{6}} = \frac{1/3}{7/6} = \frac{2}{7}. \end{aligned}$$

Table A2

Teaching and Self-teaching. ***(a)***. *The teacher's demonstration probabilities. Since the teacher's selection prior is uniform, these can be computed by taking each posterior probabilities in Table A1(b) and divide that by the corresponding value in the last row of the same table.* ***(b)***. *The teacher's instruction probabilities, obtained from summing over the y of the demonstration probabilities.* ***(c)***. *The self-teacher's selection probabilities, obtained from summing over the hypotheses of the instruction probabilities then multiply by the learner's prior.* $P_T(x=2)$ *is the highest.*

$P_T(x, y \mid h)$	h_1	h_2	h_3	h_4
$x=1, y=0$	0	$2/7$	$2/7$	$2/11$
$x=1, y=1$	$6/11$	0	0	0
$x=2, y=0$	0	0	$3/7$	$3/11$
$x=2, y=1$	$3/11$	$3/7$	0	0
$x=3, y=0$	0	0	0	$6/11$
$x=3, y=1$	$2/11$	$2/7$	$2/7$	0

(a) *Teaching with Demonstration*

$P_T(x \mid h)$	h_1	h_2	h_3	h_4
$x=1$	$6/11$	$2/7$	$2/7$	$2/11$
$x=2$	$3/11$	$3/7$	$3/7$	$3/11$
$x=3$	$2/11$	$2/7$	$2/7$	$6/11$

(b) *Teaching with instruction*

	$P_T(x)$
$x=1$	$25/77$
$x=2$	$27/77$
$x=3$	$25/77$

(c) *Self-teaching*

All the demonstration probabilities are recorded in Table A2(a).[8] Next, we compute the **teacher's instruction selection probability** according to Equation (6b) by marginalizing the teacher's demonstration distribution over y: $P_T(x|h) = \sum_{y \in \mathcal{Y}} P_T(x, y|h)$. Again, continuing our example and assuming h_2 is the true hypothesis, and the teacher wants to instruct the learner to select $x = 1$:

$$P_T(x=1|h_2) = \sum_{y \in \mathcal{Y}} P_T(x=1, y|h_2) = \frac{2}{7} + 0 = \frac{2}{7}.$$

All the teacher's instruction selection probabilities are recorded in Table A2(b). Finally, we can compute the **self-teacher's selection probability** according to Equation (8b) by marginalizing the instruction selection distribution over all possible hypotheses g: $P_T(x) = \sum_{g \in \mathcal{H}} P_T(x, |g)\, P_L(g)$. Finally, using our worked example, we can show that the self-teacher will intervene $x = 1$ with probability given by:

$$\begin{aligned} P_T(x=1) &= \sum_{g \in \mathcal{H}} P_T(x=1|g)\, P_L(g) \\ &= \frac{6}{11} \cdot \frac{1}{4} + \frac{2}{7} \cdot \frac{1}{4} + \frac{2}{7} \cdot \frac{1}{4} + \frac{2}{11} \cdot \frac{1}{4} \\ &= \frac{42 + 22 + 22 + 14}{77} \cdot \frac{1}{4} = \frac{25}{77}. \end{aligned}$$

The self-teaching selection probabilities are recorded in Table A2(c), which matches the results for the self-teaching model shown in Figure 3(a).

[8] These probabilities are not the probabilities at convergence but are the teaching probabilities in the first iteration, i.e., $P_{T_0}(x, y|h)$ as described after Equation (5).

Table A3

Calculating the predictive probability, posterior entropy and expected information gain

	$P_L(y \mid x)$	$H(h \mid x, y)$
$x = 1, y = 0$	$3/4$	$\log 3$
$x = 1, y = 1$	$1/4$	0
$x = 2, y = 0$	$1/2$	$\log 2$
$x = 2, y = 1$	$1/2$	$\log 2$
$x = 3, y = 0$	$1/4$	0
$x = 3, y = 1$	$3/4$	$\log 3$

(a) *The predictive probability and posterior entropy for all observations*

	$EIG(x)$
$x = 1$	$\log 4 - 3/4 \log 3$
$x = 2$	$\log 4 - \log 2$
$x = 3$	$\log 4 - 1/4 \log 3$

(b) *The expected information gain for each intervention*

Expected information gain.

To compute the expected information gain in Equation (4), we first compute the the **prior entropy** according to Equation (1):

$$H(h) = \sum_{h \in \mathcal{H}} P_L(h) \log \frac{1}{P_L(h)} = 4 \times \left(\frac{1}{4} \log \frac{1}{1/4} \right) = \log 4.$$

Then, we compute the **posterior entropy** according to Equation (2). For example, the posterior entropy after observing $x = 1, y = 0$ is calculated as follows:

$$\begin{aligned} H(h \mid x = 1, y = 0) &= \sum_{h \in \mathcal{H}} P_L(h \mid x = 1, y = 0) \log \frac{1}{P_L(h \mid x = 1, y = 0)} \\ &= 0 \log 0 + \frac{1}{3} \log 3 + \frac{1}{3} \log 3 + \frac{1}{3} \log 3 = \log 3. \end{aligned}$$

Then, we need to compute the **predictive probability**, $P_L(y \mid x) = \sum_{h \in \mathcal{H}} P_L(y \mid x, h)\, P_L(h)$. Continuing our example, the predictive probability of $y = 0$, given $x = 1$ is given by:

$$P_L(y = 0 \mid x = 1) = \sum_{h \in \mathcal{H}} P_L(y \mid x, h)\, P_L(h) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} = \frac{3}{4}.$$

The predictive probabilities and posterior entropy values are recorded in the first and second column of Table A3(a), respectively. Finally, using the prior entropy calculated and the values in Table A3(a), we can obtain the **expected information gain** according to Equation (4). Finally, we can calculate the expected information gain from intervening on $x = 1$ like so:

$$\begin{aligned} EIG(x = 1) &= H(h) - \sum_{y \in \mathcal{Y}} P(y \mid x = 1)\, H(h \mid x = 1, y) \\ &= \log 4 - \left(\frac{3}{4} \log 3 + \frac{1}{4} 0 \right) = \log 4 - \frac{3}{4} \log 3. \end{aligned}$$

All EIG values are recorded in Table A3(b).