

Teacher-Aware Active Robot Learning

Mattia Racca, Antti Oulasvirta and Ville Kyrki

Abstract—This paper investigates Active Robot Learning strategies that take into account the effort of the user in an interactive learning scenario. Most research claims that Active Learning’s sample efficiency can reduce training time and therefore the effort of the human teacher. We argue that the performance driven query selection of standard Active Learning can make the job of the human teacher difficult, resulting in a decrease in training quality due to slowdowns or increased error rates. We investigate this issue by proposing a learning strategy that aims to minimize the user’s workload by taking into account the flow of the questions. We compare this strategy against a standard Active Learning strategy based on uncertainty sampling and a third strategy being an hybrid of the two. After studying in simulation the validity and the behavior of these approaches, we conducted a user study where 26 subjects interacted with a NAO robot embodying the presented strategies. We reports results from both the robot’s performance and the human teacher’s perspectives, observing how the hybrid strategy represents a good compromise between learning performance and user’s experienced workload. Based on the results, we provide recommendations on the development of Active Robot Learning strategies going beyond robot’s performance.

Index Terms—Active Learning; Human-Robot Interaction; Interactive Machine Learning

I. INTRODUCTION

Service robots have the potential to become the definitive helpers in our households, hospitals and schools. Yet many challenges lie ahead and skill learning is one of them. As pre-programming robots for every situation and environment is unfeasible, service robots need to either generalize already known skills or learn new ones after deployment. We are interested in developing methods for robots to learn by interacting with their end users.

Interactive Machine Learning (IML) aims to close the gap between robots and their human teachers [1]. Several IML techniques have been proposed and used in robotics, mainly characterized by how the user contributes to the robot’s learning process. Together with other IML techniques like *Learning from Demonstration* (LfD) [2] and *Learning from Critique* (LfC) [3], [4], *Active Learning* (AL) [5] has attracted the attention of the robotics community. AL approaches enable the learning agent to query the user and decide where to concentrate its learning attempts. The querying process allows the active learner to choose *what to learn*, often by maximizing the information gain of each query. This *query efficiency* of

active learners reduces the number of labeled samples needed, therefore reducing training time and costs.

Due to their interactive nature, AL approaches must take into account the user in the loop as their performance partially depends on the interaction. A common claim championed by AL researchers is that through lowering the number of samples required to learn, the user’s effort is accordingly reduced [6]–[8]. However, to the best of our knowledge, this claim has never been supported by analysis of the workload of the human teacher. Research in *Human-Robot Interaction* (HRI) has however observed how people can find the questions posed by an active learner difficult to answer with respect to other available type of questions or with respect to the current context [9]–[13]. This difficulty can result in slower answering of the questions, distraction or errors in the training.

We want to challenge the idea that the sample efficiency of AL implies a reduced effort for the user acting as a teacher. To investigate this aspect, we study an information gathering problem where a robot has to learn *attributes* of certain *entities* by making queries to its user. We provide the learner with information about *categories* grouping the entities. This allows the learner to assume that *entities in the same category are more likely to share the same value for an attribute*. We tackle this learning problem with *uncertainty sampling* [14], an AL strategy that selects the most uncertain query based on the current model. To maximize the information gain, this strategy aims to make queries about entities that are as distant as possible from each other, given the categories. This technique acts as a strong baseline for our analysis. We hypothesize that the continuous context switching caused by these efficient queries can increase the workload of the human teachers, making them reply slower and be more prone to errors, consequently hindering the training.

We propose a learning strategy that takes into account the flow of its questions, i.e. the order in which the questions are asked. Informed by cognitive models of memory retrieval [15], our strategy builds on the concept that memory chunks require less mental effort to be retrieved if related chunks were recently recalled. The proposed strategy therefore minimizes the distance between consequent queries in order to reduce the memory effort required by the user to answer them. This strategy prefers queries targeting entities that are related to each other, minimizing the mental effort at the cost of less information gain. To further study this trade-off between information gain and user’s effort, we propose a third strategy that is an hybrid of the previous two and aims to maximize the information gain of queries while keeping the context changes as small as possible.

We compare the aforementioned three learning strategies

*The work of MR was supported by Academy of Finland, decision 314180. The work of AO has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 637991). The authors are with the School of Electrical Engineering, Aalto University, Finland (name.surname@aalto.fi).

and analyze them in simulation to assess their validity. We then run a user study, where 26 subjects interacted with a NAO robot, embodying the three AL strategies. In addition to performance measures like error rates, response times of the users and model quality, we also measure the workload of the participants through the NASA *Task Load Index* (TLX) [16] and collect subjective feedback. We observed how the flow of questions of the proposed strategies is positively perceived by the participants with respect to the standard performance-driven strategy. Results show how the perceived learning performance of the robot non-trivially influences the workload and the performance of the teacher, with the hybrid strategy exhibiting a favorable trade-off between learning performance and teacher workload. We discuss these results in the light of providing recommendations for the development of AL robots considering not only their own performance but also the human teacher’s side of the interaction.

II. RELATED WORK

Active Learning is a machine learning paradigm that allows the learning agent to choose the samples used for its training [5], [17]–[19]. While passive supervised strategies rely on labeled datasets, AL strategies depend on a user (often referred as *oracle*, *annotator* or *teacher*) to obtain labels for the chosen unlabeled samples. If the AL agent can select *informative* queries, models can be learned faster (i.e. with less labeled samples) than with passive strategies. This is particularly beneficial for problems where unlabeled samples are plenty but the labeling costs are high.

Given its sample efficiency and the interactive nature of the learning process, AL has been used for several applications in robotics and HRI. Hayes and Scassellati [20] enabled a robot to ask questions during the execution of a task, in order to obtain information about the feasibility of the next step. Similarly, Racca and Kyrki [12] proposed an AL technique for learning task models by combining LfD and queries expressed in natural language. Sadigh et al. [21] presented an application of AL in a *Inverse Reinforcement Learning* (IRL) scenario, where an autonomous car learned from users their preferred driving style by posing comparison queries. Their work was expanded in [22], where the utilized queries were enriched with features in order to better estimate the user’s preferences. Bullard et al. [23] also used different types of queries and proposed a method to choose between them in order to effectively use the budget of an AL agent.

Another line of research looks instead at the nature of the interaction with AL robots. Aspects like the nature of robot’s queries [10] and of the query selection mechanism [12], [24], and their impact on the quality of user’s answers [25], [26] have been studied. Also the balance of control between learner and teacher during the interaction [9], the effect of queries’ timing [27] and the transparency of the learning process [28] have been investigated. Cakmak et al. presented in [9] a detailed analysis of the interaction between non-expert users and an AL robot, proposing design guidelines regarding the interaction.

Work done in this area [9]–[13] observed how AL strategies could be perceived as difficult by the user due to their search for information gain that represents their strongest advantage. Few works have directly investigated this issue, mainly outside robotics. Culotta and McCallum [7] proposed an AL method that reduces both the number of samples and the difficulty of labeling those samples for the annotator. However, the query difficulty measure was engineered for a form filling scenario and the method was only tested in simulation. A similar technique was proposed in [29] for image annotation. Although their technique modeled the difficulty of annotating an image from real users, they did not study the impact of their technique with a user study. Bestick et al. [13] proposed an AL strategy that considers the ergonomic of human grasping as costs for a robot learning to perform handovers. Finally, Baldrige and Palmer [30] showed, in a language annotation scenario, how the best AL strategy can vary with the level of expertise of the annotators. However, their strategy did not directly take into account these annotators’ preferences.

Building on these results, we further investigate this issue by considering an information gathering problem where the order of the queries can be manipulated by the learner to ease the human teacher’s job. We study three different AL strategies that exhibit different behaviors regarding their flow of questions, together with different levels of performance. With a user study, we study the nature of the interaction, going beyond the learning performance to evaluate the users’ effort, their teaching performance and preferences regarding different strategies.

III. ACTIVE LEARNING USING CATEGORIES

In order to investigate different AL strategies and their effects, we chose an information gathering problem. In this section we first present the task at hand and how it can be modeled and learned online by asking questions. Second, we present three AL strategies, with three different goals in mind.

A. Incremental learning through categorical information

A learning agent has to learn the value of a certain attribute a for a set \mathcal{E} of entities. For example, a robot could learn the preferred location in your house (attribute) for different items (entities). The agent can actively gather information about a entity-attribute pair $\langle e, a \rangle$ by making a query $q_{e,a}$ to the user. With no additional information about the entities, the learner needs to make all possible queries $q_{e,a}, \forall e \in \mathcal{E}$ to learn about the attribute value of all entities.

We provide the learning agent a set of categories \mathcal{C} , characterizing the entities \mathcal{E} . For each entity in \mathcal{E} and category in \mathcal{C} , we define $w_{e,c}$ as the relevance of category c for entity e . Following from the previous example, the robot could know that items can be grouped by the function they accomplish. The learner can then make the assumption that *entities in the same category are more likely to share the same attribute value*. Following this *intra-category consistency* assumption,

we model the probability of attribute a given a category $c \in \mathcal{C}$ as

$$p(a = x|c) \sim f_{c,a}(x|\theta_{c,a}), \quad (1)$$

where $f_{c,a}(x|\theta_{c,a})$ is a distribution suiting the nature of a .

As we want the agent to learn in an interactive fashion, we need a model that can be incrementally updated after each query $q_{e,a}$. We therefore adopt a Bayesian approach with priors $p(\theta_{c,a})$ over $p(a = x|c)$. These priors can be initialized as uninformative at the beginning of training or encode available prior knowledge. After each query $q_{e,a}$, the user’s answer r gives the learner not only the value of a for entity e , but also a way to compute the posterior distribution $p(\theta_{c,a}|q_{e,a}, r)$ over the categories. Essentially, the query-answer pair acts as an observation for $p(a = r|c)$, weighted by the relevance $w_{e,c}$. Once computed, the posterior becomes the new prior and the online training can continue.

Following the intra-category consistency assumption, the learner can make predictions over the value of attribute a for entity e through the categories \mathcal{C} . The learner can estimate the probability $p(a = x|e)$ as

$$p(a = x|e) \sim f_{e,a}(x) = \sum_{c \in \mathcal{C}} \bar{w}_{e,c} f_{c,a}(x|\theta_{c,a}), \quad (2)$$

where $f_{e,a}(x)$ is a weighted mixture of $f_{c,a}(x)$ and $\bar{w}_{e,c} = w_{e,c} / \sum_c w_{e,c}$ are the normalized relevances.

This ability to make predictions allows the learner to evaluate the informativeness of a query $q_{e,a}$. If the learner can reliably estimate the answer r of a query $q_{e,a}$ with Eq. 2, it should prefer other more informative queries, where the answer is uncertain. This prepares the ground for AL techniques.

Our learning scenario: To evaluate our method, we designed a task where a robot has to learn about animals and their attributes by asking questions to a user. We use the *Animal with Attributes 2* (AwA2) dataset [31]. Used in Attribute Based Classification research [32]–[34], the dataset consists of images of 50 *classes* representing different mammals. Each class is described by 85 *semantic attributes*, based on the work of Osherson et al. [35] who collected the judgments of human subjects on the *relative strength of association* between attributes and mammals. In our scenario, we want the agent to learn these class-attribute relations by asking questions to the user (classes in AwA2 are our entities). We therefore use the dataset as a ground truth for these relations.

AwA2 does however not provide the categorical information we need. Similar to [33], [36], we obtain this information by exploiting the hierarchical representation of WordNet, a lexical database of English [37]. We use the super-subordinate relations embedded in WordNet to extract significant categories for the AwA2 entities and build a tree where entities are the leaves and categories are the non terminal nodes. Algorithm 1 summarizes the tree building process, starting from the Entity set \mathcal{E} and a desired root of the tree. On the Entity-Category tree \mathcal{T} , we compute the relevance scores $w_{c,e}$ as

$$w_{c,e} = \exp(-\gamma d(c, e)), \quad (3)$$

Algorithm 1 Build Entity-Category tree

Input: Entity set \mathcal{E} , Root of the tree \mathcal{R} , WordNet

Output: Entity-Category tree \mathcal{T} , Category set \mathcal{C}

- 1: Initialized tree \mathcal{T} with root in \mathcal{R}
 - 2: **for all** entities $e \in E$ **do**
 - 3: # find WordNet hypernym path leading from e to \mathcal{R}
 - 4: $\rho \leftarrow \{\}$; $\omega \leftarrow$ parent node of e
 - 5: **while** $\omega \neq \mathcal{R}$ **do**
 - 6: Append ω to ρ
 - 7: $\omega \leftarrow$ parent node of ω
 - 8: **end while**
 - 9: Add ρ to \mathcal{T}
 - 10: **end for**
 - 11: Prune trivial nodes from \mathcal{T} (nodes with a single child)
 - 12: $\mathcal{C} \leftarrow$ all non-terminal nodes of \mathcal{T}
-

where $d(c, e)$ is a distance metric defined as the number of edges between nodes e and c in \mathcal{T} and γ is a scale parameter.

We now adapt the previously introduced model to this task. In the following, we will concentrate on learning a single attribute a at a time. We therefore drop the subscript a to simplify the notation.

In AwA2 the attributes are boolean: an entity either has or has not a certain attribute. We hence model the probabilities $p(a = x|c)$ as Bernoulli distributions $\text{Ber}(x|\theta_c)$. Consequently, we use Beta distributions $\text{Beta}(\theta_c|\alpha_c, \beta_c)$ as the conjugate prior over the Bernoulli distributions. After each query q_e and relative answer r , we update the posterior distribution $p(\theta_c|q_e, r)$ as

$$\begin{aligned} p(\theta_c|q_e, r) &\propto p(\theta_c|\alpha_c, \beta_c)p(q_e, r|\theta_c) = \\ &= \begin{cases} \text{Beta}(\theta_c|\alpha_c + \bar{w}_{c,e}, \beta_c) & \text{if } r = \text{yes} \\ \text{Beta}(\theta_c|\alpha_c, \beta_c + \bar{w}_{c,e}) & \text{if } r = \text{no} \end{cases} \quad (4) \end{aligned}$$

i.e. each category-attribute distribution is updated with the evidence that queried entity e has or has not the attribute a , weighted with the entity-category relevance $\bar{w}_{c,e}$.

Algorithm 2 details the learning process. During the interaction, the agent can ask questions from a pool \mathcal{Q} built from the entity set \mathcal{E} and an attribute a . Examples of queries are “Do giraffes have patches?” and “Are rhinos strong?”. As long as the user is willing to answer them, the AL agent can select a new query q^* based on the strategies explained in the next section, make the query, receive an answer and integrate the answer in the model.

B. Query Selection Strategies

To learn the model presented in the previous section, a learning agent could ask questions targeting random entities and update the model with the obtained answers. To evaluate the performance of the agent, we can stop the learner at any time during the training (i.e. after any number of questions) and evaluate its predictions on the unseen entities.

However, making predictions through Eq. 2 allows the agent to estimate which queries are more informative than others. This is where AL comes into play: the learner can

Algorithm 2 Attribute Active Learning

Input: Entity-Category tree \mathcal{T} , Entity set \mathcal{E} , Attribute a
Output: Updated Category priors $p(\theta_c)$, $\langle e, a \rangle$ pairs

- 1: $\mathcal{Q} \leftarrow$ create query pool from Entity set \mathcal{E} and a
 - 2: **while** user wants to answer queries and $\mathcal{Q} \neq \{\emptyset\}$ **do**
 - 3: **for all** questions $q \in \mathcal{Q}$ **do**
 - 4: $S_q \leftarrow$ compute query score [see Section III-B]
 - 5: **end for**
 - 6: $q^* = \operatorname{argmax}_q S_q$
 - 7: $r^* \leftarrow$ make selected query q^* and wait for answer
 - 8: update model with r^* [see Eq. 4]
 - 9: remove q^* from \mathcal{Q}
 - 10: **end while**
-

ask the questions in a *certain order* so that a score is maximized. In particular, our is a case of *instance-based AL*, where the queries are the instances to be selected from the query pool [10]. Commonly in classic AL, the score to be maximized is related to the informativeness of the questions: more informative questions bring more information and should be asked earlier in order to speed up the training [19].

We study three AL strategies. First, we propose a *Classic (C)* AL strategy aiming to maximize the information gain of each query based on the current model. For this strategy we use an uncertainty sampling technique [14] that selects the query based on the entropy of the current prediction, by computing

$$s_{q,\mathbf{C}} = \mathbb{H}(\hat{f}_e(x)), \quad (5)$$

where $\hat{f}_e(x) = \operatorname{Ber}(x|\theta = \sum_c \bar{w}_{c,e}\theta_c)$ is an approximation of the full Bernoulli Mixture, as the entropy of a Bernoulli Mixture cannot be computed in close form. The entropy of a Bernoulli distribution ranges from 0 to 1, the closer to 1 the more uncertain. This score makes the learner select the query that is most uncertain about, given the current model.

As we want to study how non-performance driven AL affects the user during the interaction, we propose a *Memory Effort* strategy (**M**), inspired by the ACT-R model of declarative memory [15]. We use the concept of associative strength between memory chunks, saying that chunks of memory that are frequently associated with recently retrieved chunks have higher activation and require less effort to be retrieved [15], [38], [39]. Following this concept, strategy **M** maximizes the following score

$$s_{q,\mathbf{M}} = \exp(-\delta d(e, p)), \quad (6)$$

where $d(e, p)$ is the distance (used in Eq. 3) between the entity e target of query q and the entity p target of the *previous query* and δ being a scale parameter similar to γ . The score represents the similarity of two entities based on the tree \mathcal{T} .

These strategies produce two completely different flows of queries. Fig. 1 shows a representative flow for each strategy. A learner using strategy **C** targets entities in the tree \mathcal{T} as far from each other as possible to obtain the most information gain from them. A learner using **M** instead groups its queries using

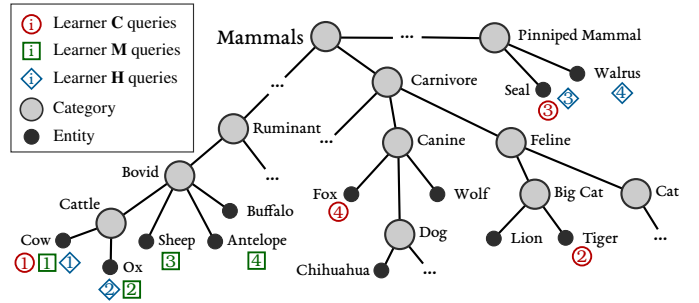


Fig. 1: Part of the Entity-Category tree \mathcal{T} and representative query flow for each learning strategy.

the structure of \mathcal{T} to reduce the memory effort of the user. Notice how strategy **M** is not optimal under the AL point of view, trading information gain by asking about closely related entities.

Finally, we present an *Hybrid* strategy (**H**), that is a combination of the **C** and the **M** strategies. The score to be maximized is defined as

$$s_{q,\mathbf{H}} = \phi s_{q,\mathbf{C}} + (1 - \phi) s_{q,\mathbf{M}}, \quad (7)$$

where parameter $\phi \in [0, 1]$ controls the trade-off between the other two strategies.

C. Simulation Experiments

We evaluated the proposed model and strategies in simulation.¹ We simulated the learning of all 85 attributes for the 50 entities of the AWA2 dataset for each learning strategy. In addition, we simulated a passive learner **P** as a baseline. Learner **P** can ask questions but does not use our category model and therefore cannot make predictions. Therefore, for learner **P** the order in which the questions are posed does not matter and they can be asked at random.

We constructed the Entity-Category tree \mathcal{T} assuming the root to be the WordNet node *Mammal*. Fig. 1 shows part of \mathcal{T} . We initialized the model of each learner with uninformative priors for $p(\theta_c|\alpha_c, \beta_c)$. Parameters γ and δ were experimentally set to 0.7. For learner **H**, we set ϕ to 0.8 in order to obtain a specific behavior regarding the query flow. As shown in Fig. 1, learner **H** would first target similar entities in a category but then switch to a unrelated category in order to gain more information.

Results: We analyzed the learning strategies under three quality measures. First, we define the predictive power $P(n)$ of a learner as the number of unseen entities (i.e. entities not yet queried) for which the attribute would have been correctly predicted if the training was stopped after n questions. Notice that the predictive power goes to zero when all questions have been asked and there is nothing left to predict. Based on the

¹Code available at github.com/MattiaRacca/TeacherAwareAL

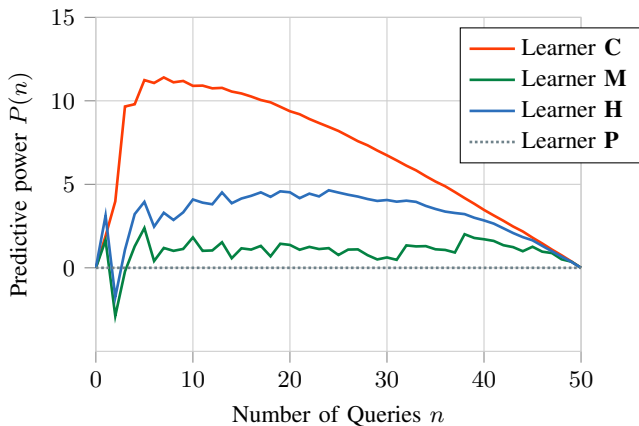


Fig. 2: Temporal evolution of the predictive power $P(n)$ of the three learning strategies relative to learner \mathbf{P} acting as baseline, averaged over the 85 attributes of AwA2.

predictive power, we define the cumulative predictive power \mathcal{P} as the sum of $P(n)$ for a complete learning session, i.e.

$$\mathcal{P} = \sum_{n=1}^{|\mathcal{E}|} P(n). \quad (8)$$

Third, we define the cumulative query similarity \mathcal{S} as the sum of the similarities between each query and the previous one (following Eq. 6) throughout a training session.

Fig. 2 shows the temporal evolution of the predictive power $P(n)$ during the training for each learner. We used learner \mathbf{P} as baseline, allowing it to predict based on a coin toss, i.e. if, after n questions, there are $|\mathcal{E}| - n$ unseen entities, learner \mathbf{P} predicts on average half of them correctly. As expected, learner \mathbf{C} performs better than \mathbf{M} and \mathbf{H} , being able on average to predict correctly the attributes of 32 unseen entities after just 5 queries, 10 more than learner \mathbf{P} 's random chance. On the other hand, learner \mathbf{M} barely performs better than the passive baseline. Learner \mathbf{H} achieves intermediate results, stemming from its trade-off nature. Fig. 3 shows the cumulative predictive power \mathcal{P} separately for each attribute in AwA2. Notice how it is possible for the learners' \mathcal{P} to be negative, i.e. worse than \mathbf{P} . This happens because the attributes in AwA2 do not always follow the intra-category consistency assumption. As an example, the attribute *domestic* conflicts with this assumption, with e.g. dogs being in the same category of wolves and foxes, and cats being closely related to lions and tigers.

Finally, Table I shows the cumulative query similarity \mathcal{S} for each learner. Learner \mathbf{M} achieves the highest \mathcal{S} , closely followed by learner \mathbf{H} . Learner \mathbf{C} has, together with the passive learner, lower \mathcal{S} scores. As expected, the \mathcal{S} scores reflect the nature of the different query selection strategies as explained in Section III-B and depicted in Fig. 1.

IV. USER STUDY

In order to study how different AL strategies impact the human teacher during the interaction with a learning robot, we

ran first a pilot study (6 participants) followed by a user study with 26 participants. The participants interacted with Nemo, a NAO robot embodying the proposed learning strategies (\mathbf{C} , \mathbf{M} and \mathbf{H}). The participants acted as teachers for the robot that was learning 6 different attributes (2 attributes per strategy) for the animals of the dataset. We investigated how the different strategies impact the speed and the quality of the teaching of the participants, recording the response times, error rates and predictive power. Furthermore, we administered the NASA-TLX questionnaire for measuring the participants' workload and a custom session questionnaire, targeting specific aspects of the proposed strategies.

Experimental Setup: Experimental setup is shown in Fig. 4. The NAO robot Nemo sits in front of the participant, close to a keyboard and a screen. The learning software ran on an external laptop, connected to the robot through ROS [40]. Nemo used animated text to speech to ask questions and express other utterances. Additionally, the questions were showed on the screen. The participants replied to the robot's question with *Yes/No/I don't know* by pressing the arrow keys on the keyboard. We opted for using a keyboard instead of more sophisticated techniques to more reliably observe the possibly small differences in response times.

Participants: Twenty-six participants (age $M = 27$, $SD = 6$, female 76%) were recruited in a university campus. Participants had different education levels (9 high school diploma, 10 bachelor's degrees, 6 master's degrees and 1 PhD), with 2 participants having a computer science or engineering background and none having experience with NAO robots. The participants were rewarded with a movie ticket as experiment incentive, together with a debriefing pamphlet.

Conditions and Protocol: Each participant interacted with all three learning strategies. The order of strategies was counterbalanced and all orderings happened at least 4 times. We used the same parameters (γ , δ , ϕ and uninformative priors) used in the simulations. Experiments lasted on average 40 minutes.

Each participant was first introduced to Nemo and received instructions for the task. A training session followed, where Nemo asked the participant whether 14 animals were mammals or not (none of them from AwA2), in order to familiarize with the answering system and the robot's voice. Each participant then engaged in three learning sessions, each

TABLE I: Cumulative Query Similarity \mathcal{S} ($M \pm SD$)

	Learner \mathbf{C}	Learner \mathbf{M}	Learner \mathbf{H}	Learner \mathbf{P}
\mathcal{S}	2.61 ± 0.66	8.24 ± 0.04	7.62 ± 0.40	1.48 ± 0.32

TABLE II: Experiment design and selected attributes.

	Session 1	Session 2	Session 3
Visual attributes (physical features)	Do ___ have paws?	Do ___ have horns?	Do ___ have claws?
Non-visual attributes (diet)	Do ___ prefer to eat fish?	Do ___ prefer to eat meat?	Are ___ herbivore?

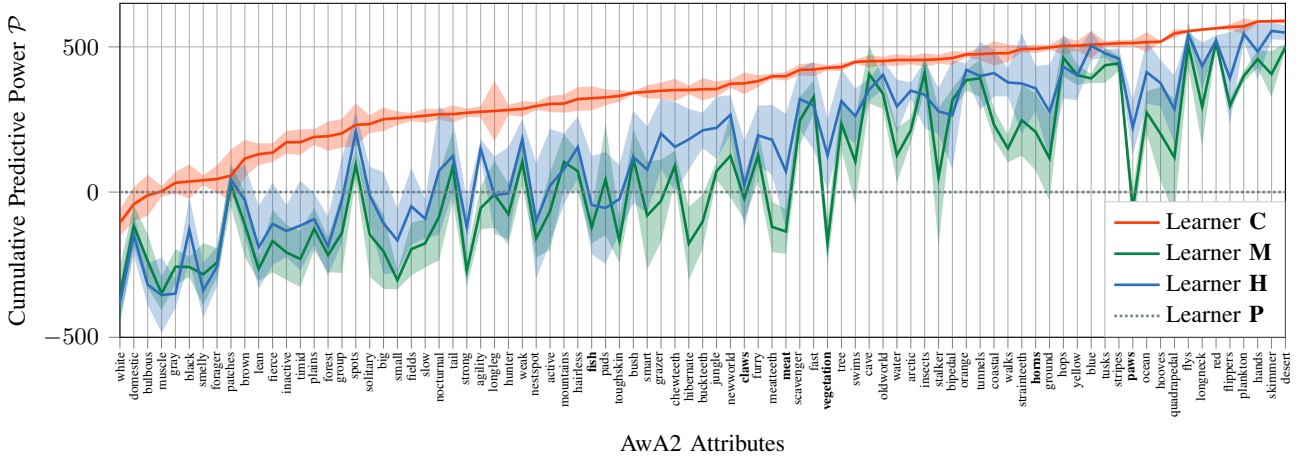


Fig. 3: Cumulative predictive power \mathcal{P} of the three strategies for each attribute in AwA2, averaged over 5 different runs (1 standard deviation highlighted by the shaded area). Attributes are sorted with respect to learner C’s performance. Attributes used in the user study are presented in bold. Best viewed in color.

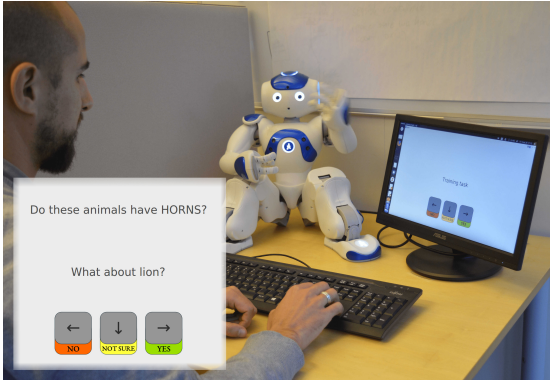


Fig. 4: Experimental setup and example of question, as shown on the screen.

interleaved by a session questionnaire and a NASA-TLX questionnaire. Each session was separated in two: in each half the robot asked questions about one of two attributes. We chose 6 attributes from the 85 available in AwA2 based on the following constraints. First, we selected attributes that have a ratio of *yes* and *no* answers over the dataset close to 0.5. We enforced this requirement to make the interaction less tedious for the participants (e.g. with a ratio close to 0 or 1, the correct answer to almost all questions is either *No* or *Yes*). Second, we selected attributes where the intra-category consistency assumption is respected, avoiding attributes from the first third of Fig. 3. Third, we wanted the attributes to be of two types, so that each strategy has one attribute per type. We therefore settled for the 3 visual attributes (physical features) and the 3 non-visual attributes (diet) shown in Table II.

For each attribute in a session, the participants had 40 seconds to answer as many questions as possible. The budget of 40 seconds was chosen after the pilot study, to allow the participants to answer on average 15 questions - a number of questions sufficient to observe differences in performance

and question flow of the three strategies. The participants were instructed to balance the time constraint and the answer quality, suggesting to reply *I don’t know* when needed. The *I don’t know* answer did not trigger any model update. At the beginning of each attribute session, the robot would announce the general question, e.g. “*Do these animals have paws?*”, and then proceed to ask about the individual animals, e.g. “*What about animal_1?*” and “*How about animal_2?*”. After asking the question, the robot would wait for the answer and record as response time the period between the end of the question and the answer. When no time budget was left, the robot would close the attribute session.

At the end of each session, Nemo evaluated its learning performance by testing its predictions on the unseen entities of the dataset. Nemo computed its current predictive power and reported it in the form of a percentage, to allow the participants to evaluate its performance.

Data Logged: After each learning session, participants filled the NASA-TLX questionnaire. We used RAW NASA-TLX as we were interested in its subscales and it is simpler to administer [16]. The participants also filled a session questionnaire with the following 3 Likert statements [1 *completely disagree* - 7 *completely agree*] about a strategy:

- 1) *The flow of Nemo’s questions felt natural*
- 2) *Nemo’s strategy made my job as teacher easy*
- 3) *Nemo’s strategy was good for its learning.*

All these statements included a *Why? Please explain* optional question. We additionally logged the queries chosen by each strategy, the participants’ answers and response time. At the end of the experiment, we collected participants’ preferences and open-ended suggestions regarding the three learning strategies.

A. Results

Tables III and IV summarize respectively the quantitative data (response times, error rates and performance measures)

TABLE III: Quantitative results: mean over session, test statistics and p -values of comparison across strategies (Friedman) and pair-wise comparison (Wilcoxon signed-rank).

	C	M	H	Friedman test	C vs M	C vs H	H vs M
Response time (RT) [s]	0.85	0.90	0.73	$\chi^2(2)=3.87, p>.05$	–	–	–
RT with visual attributes [s]	0.80	0.86	0.70	$\chi^2(2)=0.58, p>.05$	–	–	–
RT with non-visual attributes [s]	0.91	0.96	0.78	$\chi^2(2)=10.33, p<.01^{**}$	T=155.0, $p>.05$	T=71.0, $p<.05^*$	T=80.0, $p<.05^*$
Error Rate	21.4%	19.5%	11.4%	$\chi^2(2)=9.00, p<.05^*$	T=10.0, $p>.05$	T=0.0, $p<.05^*$	T=1.0, $p<.05^*$
Prediction Percentage	81.0%	51.4%	74.1%	$\chi^2(2)=48.86, p<.01^{**}$	T=17.0, $p<.01^{**}$	T=257.5, $p<.01^{**}$	T=95.5, $p<.01^{**}$
Prediction Percentage with oracles	89.4%	55.6%	77.5%	$\chi^2(2)=42.25, p<.01^{**}$	T=0.0, $p<.01^{**}$	T=1.0, $p<.01^{**}$	T=6.0, $p<.01^{**}$

TABLE IV: NASA-TLX: descriptive statistics (median), test statistics and p -values of comparison across strategies (Friedman) and pair-wise comparisons (Wilcoxon signed-rank).

	C	M	H	Friedman test	C vs M	C vs H	H vs M
Mental Demand	6.5	10.5	6.0	$\chi^2(2)=4.06, p>.05$	–	–	–
Physical Demand	1.0	1.0	1.0	$\chi^2(2)=2.60, p>.05$	–	–	–
Temporal Demand	6.0	5.5	5.0	$\chi^2(2)=1.26, p>.05$	–	–	–
Performance	6.5	9.5	5.0	$\chi^2(2)=6.89, p<.05^*$	T=100, $p>.05$	T=75.5, $p>.05$	T=36.5, $p<.01^{**}$
Effort	6.5	8.5	6.0	$\chi^2(2)=4.14, p>.05$	–	–	–
Frustration	4.0	5.0	5.0	$\chi^2(2)=6.84, p<.05^*$	T=38.5, $p>.05$	T=69.5, $p>.05$	T=10.5, $p<.01^{**}$

and the scores from the NASA-TLX questionnaire. We tested for normality with the Shapiro-Wilcox test, rejecting the null-hypothesis ($p<.05$) for each condition and score. We therefore ran the non-parametric Friedman test for differences between learners and Wilcoxon signed-rank test for pair-wise comparisons, with the participants as matching factor between samples. The error rates were computed using the participants’ answers and the dataset as ground truth.

Our hypotheses entering the user study were that learner **M** would make the participants reply (a) faster and (b) with less errors compared to learner **C**, with learner **H** obtaining intermediate results according to its hybrid nature. Much to our surprise, the participants’ response time and error rates for the different learners did not follow these expectations.

Learner **H** had the shortest response times, with an average of 0.73 s against the 0.85 s and the 0.90 s of learners **C** and **M** respectively, although statistically significant differences were observed between groups only for the case of non-visual attributes. The participants found the questions about non-visual attributes more difficult to answer with respect to the visual ones. The higher response times observed in the non-visual case reflects this perceived difficulty and the more marked differences observed between the strategies.

Regarding the error rates, participants replied incorrectly to only 11.4% of the questions with learner **H**, which is approximately half of the errors done by the participants while interacting with learners **C** and **M** (21.4% and 19.5% respectively). The participants’ error rate did not prevent learner **C** from obtaining the highest prediction percentage, with an average of 81.0% correct predictions. Learners **H** and **M** follow with 74.1% and 51.4% correct predictions respectively. Table III shows the prediction percentage in the hypothetical case of infallible users. Although the learning performance ranking does not change, we can see how learner **C** lost the most performance (8.4%) due to the users’ errors.

B. Discussion

In order to interpret the quantitative results, we analyzed the participants’ optional feedback together with the questionnaires’ scores shown in Table IV.

1) *Flow of the questions:* Although no difference was observed on the *Question Flow* score (median of 6 for all learners), 11 participants reported in the optional comments how learner **C** seemed to ask random questions and how this made the teaching stressful (4 subjects), unpredictable (1 subject) and requiring more thinking (6 subjects). Such comments suggest how the efficiency of learner **C**’s questions may have made the users’ job as teachers more challenging, causing the observed slower response time and the higher error rate with respect to learner **H**. On the other hand, learner **M** was spotted to make use of animal categories to group its queries (12 subjects), making the flow seem natural (8 subjects) and the questions easier to answer (8 subjects).

Regarding the *Good for Teacher* score, learner **C** was rated slightly lower by the participants, with a median of 5.5 against the 6 of learner **M** and **H**. Another recurrent comment about learner **M** (related to the *Good for Teacher* score) was that participants took advantage of its question flow to anticipate the questions and their answers (“*Thanks to the order of questions, a previous answer could be often applied again to the following question*” and “*As similar questions followed each other I didn’t need to think about every question separately.*”). This may have caused the participants to engage a sort of *autopilot* or simply be bored by too similar questions, lowering their attention during the task and causing the unexpectedly high error rate observed. While our model concentrated on the retrieval of information from declarative memory, the results suggest that active learners taking into account also the user’s attention may lead to more effective strategies.

Comments about the question flow of learner **H** followed a pattern similar to learner **M**, with 13 participants praising the

easiness of teaching caused by the grouping of questions. Although some participants commented about the anticipation of future questions also for learner **H**, one participant commented how the less monotonous query flow made him more attentive (“*The flow seemed natural but the slight variation of questions kept me more awake*”). This is a possible explanation for the lowest error rate and the fastest response time observed with learner **H**. A similar observation was made in [24], where a robot allowed to ask off-topic questions was perceived as *fun* and *usable* by the users.

2) *Perceived learning performance*: According to the *Good for Performance* score, the participants were able to spot the non-optimality of learner **M** with a median of 4 against the 6 and the 5.5 of learners **C** and **H** respectively. Most of the participants saw the prediction score that Nemo was providing them as their main tool to assess the robot’s performance. However, some participants tried to explain the nature of different strategies. The comments on the performance of learner **M** were particularly harsh, especially if the participant experienced other learners first (“*Better to ask questions randomly*”, “*Not the ideal strategy considering the time constraint*”). Learner **C** was in general praised for its performance (10 subjects), although some participants mentioned again how this strategy might have caused them to make mistakes (3 subjects). Interestingly, only one participant commented about the non-optimality of learner **H**. We think that this lack of negative comments was caused by learner **H**’s prediction percentages, that were often close to **C**’s ones, especially if compared to learner **M**.

We think that the perceived performance impacted not only the error rates and the response times but also how the participants perceived the workload imposed by the teaching task. Learner **M** obtained the highest (i.e. worst) scores for *Mental Demand*, *Performance* and *Effort*, although statistically significant differences were observed only for the *Performance* score compared to learner **H**. The same scores for learner **C** were on average lower and in contrast with the comments on the *Question Flow* score presented before. We think that the good performance of **C** might have made the participants pay less attention to the workload needed to achieve such performance: in a sense, the participants saw their effort pay off with learner **C**. Such thing did not happen for learner **M** and some participants reported being frustrated by it (“*As a teacher, I felt frustrated by its choice of questions and probably made mistakes due to that.*”) or willing to have more control over its queries (“*I would like to pick the questions for the robot*”). This observation supports the design recommendation made in [9]: when designing active learners, avoid uninformative queries as they could weaken the teacher’s trust in the utility of answering them.

Even though our study design did not address the effects of robot sociality on the interaction, the participants gave interesting feedback on the matter. Regardless of the learning strategy, participants reported feeling empathy for the robot, wanting to teach Nemo the best they can due to its child-like appearance and voice (8 subjects). Three participants even felt

responsible for the bad performances of learner **M**. Further research is however needed to better understand how sociality and embodiment impact the performance of learning robots compared to other learning agents like e.g. computers, building on more general results from [41], [42].

Finally, the influence of performance was also seen when participants expressed their preference regarding the three learning strategies at the end of the experiment. Twelve participants preferred learner **C**, backing their choice with comments praising its learning efficiency. Learner **H** was chosen by 10 participants, with comments concentrating on the easiness of teaching derived by the questions being grouped. Learner **M** obtained only 4 preferences. This wide range of preferences suggests how the *best strategy* is likely to be user-dependent, based e.g. on their patience or their teaching skills as observed also in [9], [12], [28], [43].

V. CONCLUSIONS

In this work, we challenged the common idea that links the sample efficiency of active learners with the reduced effort for the interacting user. We proposed a novel AL strategy that takes into account the flow of the questions posed to the user, in order to minimize the mental effort of the teacher. We compared this strategy against a classic active learning strategy and a third strategy being an hybrid of the two. We studied with a user study how these different strategies affect the human teacher in an human-robot interactive learning scenario.

Our results show how the exploratory nature of active learners, although efficient, can cause problems for the human teachers. We observed how the participants felt stressed and more prone to errors when teaching to learner **C** and how this raised both their error rates and response times. These issues have been observed before [9], [11] and have their roots in the often unrecognized or underestimated fact that human teachers are not oracles and can make mistakes.

The results from the user study show also that performance cannot be neglected or traded completely for the easiness of the process, an observation in line with what is suggested in [9]. Participants of our study realized how the strategy **M** was grouping its questions and how this feature made their job as a teacher easier. However, this strategy did not take into account other interaction aspects such as the frustration of the teachers or their concentration, and how this non-trivially influenced their view of the learning system.

In our experiment, the limitations of both strategies were overcome by the hybrid strategy **H**, which combined the easiness of teaching of strategy **M** with the performance of strategy **C**. However, the preferences of the participants were far from being unanimous. As the best teaching strategy seems to be a personal choice, we think that learning strategies that can adjust their behavior based on the user’s feedback online during the interaction have great potential. Another challenge is to better understand the factors (such as the observed frustration) that emerge when we consider teaching not only as a maximal gathering of information but as a social process between a learning robot and a human teacher.

REFERENCES

- [1] S. Chernova and A. L. Thomaz, "Robot learning from human teachers," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 8, no. 3, pp. 1–121, 2014.
- [2] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and autonomous systems*, vol. 57, no. 5, pp. 469–483, 2009.
- [3] D. B. Argall, B. Browning, and M. Veloso, "Learning by demonstration with critique from a human teacher," in *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, pp. 57–64, IEEE, 2007.
- [4] B. W. Knox, P. Stone, and C. Breazeal, "Training a robot via human feedback: A case study," in *International Conference on Social Robotics*, pp. 460–470, Springer, 2013.
- [5] B. Settles, "Active learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 6, no. 1, pp. 1–114, 2012.
- [6] E. Lughofer, "Hybrid active learning for reducing the annotation effort of operators in classification systems," *Pattern Recognition*, vol. 45, pp. 884–896, Feb. 2012.
- [7] A. Culotta and A. McCallum, "Reducing labeling effort for structured prediction tasks," in *AAAI*, vol. 5, pp. 746–751, 2005.
- [8] P. Daece, T. Peltola, M. Soare, and S. Kaski, "Knowledge elicitation via sequential probabilistic inference for high-dimensional prediction," *Machine Learning*, vol. 106, pp. 1599–1620, Oct. 2017.
- [9] M. Cakmak, C. Chao, and A. L. Thomaz, "Designing interactions for robot active learners," *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 2, pp. 108–118, 2010.
- [10] M. Cakmak and A. L. Thomaz, "Designing robot learners that ask good questions," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pp. 17–24, ACM, 2012.
- [11] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza, "Power to the people: The role of humans in interactive machine learning," *AI Magazine*, vol. 35, no. 4, pp. 105–120, 2014.
- [12] M. Racca and V. Kyrki, "Active robot learning for temporal task models," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 123–131, ACM, 2018.
- [13] A. Bestick, R. Pandya, R. Bajcsy, and A. D. Dragan, "Learning human ergonomic preferences for handovers," in *Robotics and Automation (ICRA), 2018 IEEE International Conference on*, pp. 3257–3264, IEEE, 2018.
- [14] D. D. Lewis and J. Catlett, "Heterogeneous uncertainty sampling for supervised learning," in *Machine Learning Proceedings 1994*, pp. 148–156, Elsevier, 1994.
- [15] J. R. Anderson, D. Bothell, C. Lebiere, and M. Matessa, "An Integrated Theory of List Memory," *Journal of Memory and Language*, vol. 38, pp. 341–380, May 1998.
- [16] S. G. Hart, "Nasa-task load index (NASA-TLX); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, pp. 904–908, Sage Publications Sage CA: Los Angeles, CA, 2006.
- [17] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *Journal of artificial intelligence research*, vol. 4, no. 1, pp. 129–145, 1996.
- [18] D. Angluin, "Queries and concept learning," *Machine learning*, vol. 2, no. 4, pp. 319–342, 1988.
- [19] Y. Fu, X. Zhu, and B. Li, "A survey on instance selection for active learning," *Knowledge and information systems*, vol. 35, no. 2, pp. 249–283, 2013.
- [20] B. Hayes and B. Scassellati, "Discovering task constraints through observation and active learning," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pp. 4442–4449, IEEE, 2014.
- [21] D. Sadigh, A. Dragan, S. S. Sastry, and S. A. Seshia, "Active preference-based learning of reward functions," in *Robotics: Science and Systems*, 2017.
- [22] C. Basu, M. Singhal, and A. D. Dragan, "Learning from richer human guidance: Augmenting comparison-based learning with feature queries," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 132–140, ACM, 2018.
- [23] K. Bullard, A. L. Thomaz, and S. Chernova, "Towards Intelligent Arbitration of Diverse Active Learning Queries," in *Intelligent Robots and Systems (IROS), 2018 IEEE/RSJ International Conference on*, pp. 6049–6056, IEEE, 2018.
- [24] J. Thomason, A. Padmakumar, J. Sinapov, J. Hart, P. Stone, and R. J. Mooney, "Opportunistic active learning for grounding natural language descriptions," in *Conference on Robot Learning*, pp. 67–76, 2017.
- [25] S. Rosenthal, M. Veloso, and A. K. Dey, "Acquiring accurate human responses to robots' questions," *International Journal of Social Robotics*, vol. 4, no. 2, pp. 117–129, 2012.
- [26] V. Gonzalez-Pacheco, M. Malfaz, A. Castro-Gonzalez, J. C. Castillo, F. Alonso, and M. A. Salichs, "Analyzing the Impact of Different Feature Queries in Active Learning for Social Robots," *International Journal of Social Robotics*, vol. 10, no. 2, pp. 251–264, 2018.
- [27] J. Baraglia, M. Cakmak, Y. Nagai, R. Rao, and M. Asada, "Initiative in robot assistance during collaborative task execution," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, pp. 67–74, IEEE Press, 2016.
- [28] C. Chao, M. Cakmak, and A. L. Thomaz, "Transparent active learning for robots," in *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pp. 317–324, IEEE, 2010.
- [29] S. Vijayanarasimhan and K. Grauman, "What's it going to cost you?: Predicting effort vs. informativeness for multi-label image annotations," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 2262–2269, IEEE, 2009.
- [30] J. Baldrige and A. Palmer, "How well does active learning actually work?: Time-based evaluation of cost-reduction strategies for language documentation," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pp. 296–305, Association for Computational Linguistics, 2009.
- [31] Y. Xian, C. H. Lampert, B. Schiele, and Z. Akata, "Zero-shot learning - a comprehensive evaluation of the good, the bad and the ugly," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 07 2018.
- [32] C. H. Lampert, H. Nickisch, and S. Harmeling, "Attribute-Based Classification for Zero-Shot Visual Object Categorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, pp. 453–465, Mar. 2014.
- [33] Z. Al-Halah and R. Stiefelhagen, "How to transfer? zero-shot object recognition via hierarchical transfer of semantic attributes," in *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*, pp. 837–843, IEEE, 2015.
- [34] Z. Al-Halah, M. Tapaswi, and R. Stiefelhagen, "Recovering the missing link: Predicting class-attribute associations for unsupervised zero-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5975–5984, 2016.
- [35] D. N. Osherson, J. Stern, O. Wilkie, M. Stob, and E. E. Smith, "Default probability," *Cognitive Science*, vol. 15, no. 2, pp. 251–269, 1991.
- [36] M. Rohrbach, M. Stark, and B. Schiele, "Evaluating knowledge transfer and zero-shot learning in a large-scale setting," in *2011 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1641–1648, June 2011.
- [37] G. A. Miller, "WordNet: a lexical database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.
- [38] J. R. Anderson, *How can the human mind occur in the physical universe?* Oxford University Press, 2009.
- [39] S. Jo, R. Myung, and D. Yoon, "Quantitative prediction of mental workload with the ACT-R cognitive architecture," *International Journal of Industrial Ergonomics*, vol. 42, pp. 359–370, July 2012.
- [40] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, p. 5, Kobe, 2009.
- [41] A. Powers, S. Kiesler, S. Fussell, and C. Torrey, "Comparing a computer agent with a humanoid robot," in *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, pp. 145–152, IEEE, 2007.
- [42] W. A. Bainbridge, J. W. Hart, E. S. Kim, and B. Scassellati, "The benefits of interactions with physically present robots over video-displayed agents," *International Journal of Social Robotics*, vol. 3, no. 1, pp. 41–52, 2011.
- [43] M. Cakmak and A. L. Thomaz, "Optimality of human teachers for robot learners," in *Development and Learning (ICDL), 2010 IEEE 9th International Conference on*, pp. 64–69, IEEE, 2010.