Interaction Considerations in Learning from Humans

Pallavi Koppol* , **Henny Admoni** , **Reid Simmons**
Carnegie Mellon University
{pkoppol, hadmoni, rsimmons}@andrew.cmu.edu

Abstract

The ability to learn from large quantities of complex data has led to the development of intelligent agents such as self-driving cars and assistive devices. This data often comes from people via interactions such as labeling, providing rewards and punishments, and giving demonstrations or critiques. However, people's ability to provide high-quality data can be affected by human factors of an interaction, such as induced cognitive load and perceived usability. We show that these human factors differ significantly between interaction types. We first formalize interactions as a Markov Decision Process, and construct a taxonomy of these interactions to identify four archetypes: *Showing, Categorizing, Sorting,* and *Evaluating*. We then run a user study across two task domains. Our findings show that *Evaluating* interactions are more cognitively loading and less usable than the others, and *Categorizing* and *Showing* interactions are the least cognitively loading and most usable.

1 Introduction

Intelligent agents such as self-driving cars, recommendation engines, and assistive devices are becoming fixtures in everyday society due to their growing ability to learn from large-scale data and to personalize based on data from individuals. Approaches to collecting data from people include asking for annotations on video and images [Real *et al.*, 2017], ratings of behavior [Daniel *et al.*, 2015], task demonstrations [Abbeel and Ng, 2004], critiques or corrections of proposed trajectories [Cui and Niekum, 2018; Bajcsy *et al.*, 2018], and preferences between options [Sadigh *et al.*, 2017]. Distinctions between these techniques have led to a growing body of work on understanding them relative to each other. Different interaction types can be used in combination to accelerate learning [Palan *et al.*, 2019], better leverage people as teachers [Bullard *et al.*, 2018], and differ in the amount of implicit information they encode [Jeon *et al.*, 2020]. Learning interactions are often selected based on how informative they are for a learner, without examining how human teachers may differently perceive and respond to those interactions. However, people's experiences with different interactions can affect the data they provide.

In order to collect high-quality data, whether via active learning or curated training sets, i.e. passive learning, it is necessary to leverage data collection processes that accommodate people's limitations. People provide noisy data, are biased towards providing positive rewards, and get fatigued [Amershi *et al.*, 2014]. Several of the shortcomings in people's teaching capabilities may relate to the fact that as the cognitive load (i.e. the portion of working memory being utilized) on an individual increases, they grow more easily distracted and have worse task performance [Sweller, 1988]. Interaction design can be used to modulate cognitive load in human learners [Chandler and Sweller, 1991]; that is, the way a task is presented can affect how burdensome it is and may ultimately affect the quality of the data it produces.

Our key insights are twofold. First, that *we can formalize interactions as a Markov Decision Process (MDP) and taxonomize them based on how data is provided to and acted upon by human teachers*. This taxonomy allows us to analyze groups of similar interactions. It also enables us to empirically evaluate our second insight: *these groups of interaction types result in differences in human factors such as cognitive load*. Such factors are important to characterize, as they may lead to downstream data quality effects.

We present a model-agnostic MDP to formalize interactions, and a taxonomy of interaction types for human-agent learning using four features: the amount of data the user is asked to respond to, the amount of the data the user provides in their response, the granularity of the user's response, and the responses the user can choose from. We analyze current paradigms for learning from people, and find four distinct interaction archetypes: *Showing*, *Categorizing*, *Sorting*, and *Evaluating*. We also present results from a user study designed to identify differences between interaction types in terms of human factors related to data quality, such as cognitive load, confidence, and subjective usability. We find that *Evaluating* interaction types, where people identify good behavior, are the most cognitively loading and least usable in both of the study's task domains. *Categorizing* (i.e., assigning a positive or negative reward) and *Showing* (i.e., giving demonstrations) are less cognitively loading and more usable.

*Contact Author

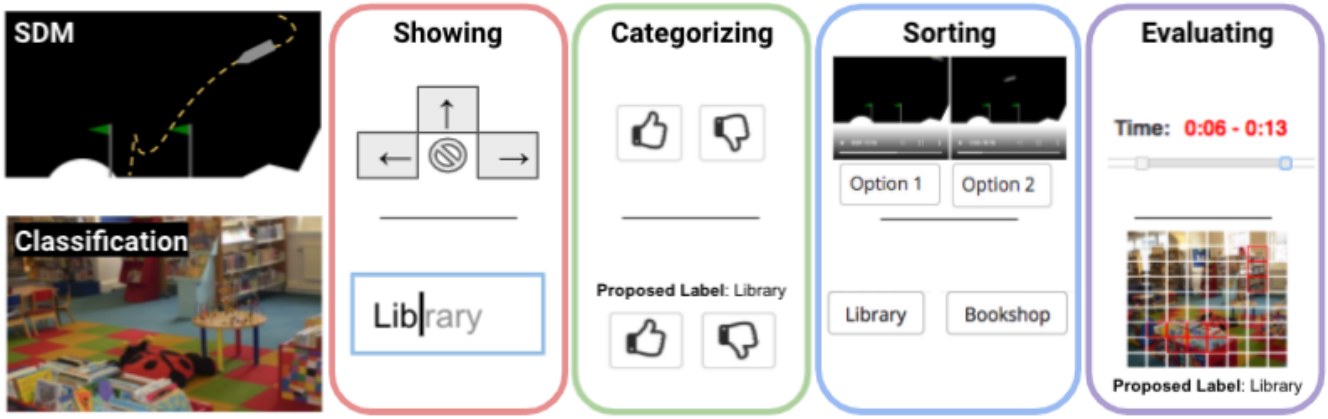


Figure 1: We identify and evaluate four interaction archetypes for sequential decision making (SDM, top) and Classification (bottom) tasks.

2 Related Work

Quality control research for data collection investigates how users, often crowd workers, can provide good data via incentives and task design [Lease, 2011; Kittur *et al.*, 2013]. Our research differs in that we study how human factors, such as cognitive load and usability, differ between various interaction types. The correlations between cognitive load, usability, and task performance (e.g. providing good data) have been observed and studied throughout cognitive psychology [Sweller, 1988], human-computer [Longo, 2018] and human-robot interaction [Prewett *et al.*, 2010].

Researchers categorized the questions people ask while learning into three interaction types: labels, demonstrations, and feature queries [Cakmak and Thomaz, 2012]. Additional research into developing learning agents that emulate the rich learning interactions people use has involved combining multiple interaction types [Bullard *et al.*, 2018] to better leverage human teachers, leveraging trade-offs between interactions [Palan *et al.*, 2019], and exploring explicit and implicit information transfer [Jeon *et al.*, 2020].

Our research extends this body of work by providing a general taxonomy of interaction types as well as a principled framework for representing model-agnostic interactions. Based on our taxonomy (justified in Section 3), we discuss four categories of interactions.

Showing. Inverse reinforcement learning (IRL) is a learning from demonstration technique for recovering a reward function from which to train a policy [Ng *et al.*, 2000; Abbeel and Ng, 2004]. Behavioral cloning learns a policy directly [Bain and Sammut, 1995; Schaal, 1999]. While demonstrations can be highly informative, people are limited in the number of examples they can provide and by their expertise.

Categorizing. This type of learning is commonly used for classification and regression. For example, computer vision leverages popular large-scale labeled datasets [Deng *et al.*, 2009; Everingham *et al.*, 2010; Lin *et al.*, 2014]. Labels also include the assignment of rewards to actions, as in bandit problems and reinforcement learning with human feedback [Kober *et al.*, 2013; Daniel *et al.*, 2015]. The informativeness of labels is limited by the size of the label set, and people are known to give both overly positive rewards [Amershi *et al.*, 2014] and shifting ratings [O'Connor and Cheema, 2018].

Sorting. Preference elicitation is an active area of research, especially in recommendation engines [Fürnkranz and Hüllermeier, 2011]. Comparison and ranking-based approaches for learning reward functions are increasingly common [Sadigh *et al.*, 2017; Bıyık *et al.*, 2020; Wirth *et al.*, 2017]. These interactions are precise, and thought to be low user effort. The technique is good at fine-tuning, but the information that can be gained from each query is limited.

Evaluating. Corrections are feedback on a proposed set of actions either during or after task execution. These can be physical or simulated [Bajcsy *et al.*, 2018; Jain *et al.*, 2015]. Users can also mark good or bad regions of a trajectory via critiques [Cui and Niekum, 2018]. Credit-assignment interactions, such as the one posed in [Jeon *et al.*, 2020], can be construed as a form of critique where the user is limited to identifying only one good region. Off-switch games can be considered as another special case of critiques where trajectories are segmented into a singular allowed section preceding a singular disallowed section [Hadfield-Menell *et al.*, 2016].

3 Taxonomy of Interaction Types

Many interaction types share fundamental similarities in how people interface with them. By taxonomizing the space of interaction types along those lines, we can more tractably compare clusters of interactions. We construct a novel taxonomy that characterizes interaction types along four dimensions: the action batch size of the learner's queries, the action batch size of the user's responses, the number of intervention opportunities available to the user per query, and the number of response choices available for a user to select from per query. We identify four interaction archetypes, termed: *Showing, Categorizing, Sorting, and Evaluating*.

3.1 Representing Model-Agnostic Interactions

In order to represent interactions, we must first introduce some terminology. Let a *query* q refer to data that the user is prompted to respond to. An *annotation* is the feedback that the user gives in response to a query. An *interaction type* is

Interactions	Query Size	Response Size	Intervention Options	Response Choice Space	References
Showing					
Demonstrations	0	T	0	$\|A_L\|^T$	[Ng *et al.*, 2000; Abbeel and Ng, 2004; Ramachandran and Amir, 2007; Ziebart *et al.*, 2008; Bain and Sammut, 1995; Schaal, 1999]
Categorizing					
Labels	T	1	0	$\|A_U\|$	[Deng *et al.*, 2009; Everingham *et al.*, 2010; Lin *et al.*, 2014]
Reward & Punishment	T	1	0	$\|\{-1,+1\}\|$	[Kober *et al.*, 2013; Daniel *et al.*, 2015]
Sorting					
Rankings	$T \cdot N$	N	0	$N!$	[Fürnkranz and Hüllermeier, 2011; Wirth *et al.*, 2017; Bıyık *et al.*, 2020]
Preferences	$T \cdot 2$	2	0	2!	[Sadigh *et al.*, 2017]
Evaluating					
Corrections	T	$0 \leq i \leq T$	2^T	$\|A_U\|^i$	[Bajcsy *et al.*, 2018; Jain *et al.*, 2015]
Critiques	T	$0 \leq i \leq T$	2^T	2^T	[Cui and Niekum, 2018; Jeon *et al.*, 2020; Hadfield-Menell *et al.*, 2016]

Table 1: We divide interactions into four clusters. T is the finite time horizon of a presented trajectory (1 in one-shot instances), N is the number of trajectories of length T in a query or response, A_U is the set of user actions, A_L is the set of learner actions, and i is a subset of T.

the format of a query (i.e., "Showing", "Categorizing", "Sorting", or "Evaluating"). An *interaction instance* is a specific query, e.g. "Should the label be 'library' or 'bookshop'?" Finally, an *interaction session* is a series of interaction instances of a particular interaction type, and associated annotations.

We choose to model interaction sessions as Markov Decision Processes (MDPs) for several reasons. MDPs provide a sequential decision-making paradigm that captures how, in active and passive learning, people provide a series of annotations over the course of an interaction session. Furthermore, with this paradigm, we can treat human teachers as agents making decisions over their own action and state spaces, rather than as oracles in possession of data that is always equally accessible. This allows us to account for imperfect decision-making due to human factors (i.e. cognitive load, usability). Finally, this enables us to analyze interaction types separately from any underlying learning models. The interaction type is a means to obtain data, and the learning model (e.g. Gaussian process, neural network, Q-learning) consumes that data. This distinction enables us to discuss interaction types in terms of the user's actions and the learner's actions, to analyze both passive and active data collection, and to assess interactions regardless of learning objectives.

We subsequently define this MDP in further detail. First, we define a user U as an agent interacting with a learner L via queries. The interactions between U and L can be situated in passive or active learning contexts. Let A_L define the set of actions available to the learner L and $a_t \in A_L$ the action taken at time t. For example, in an autonomous driving task, A_L consists of available steering controls. Let $s_t \in S_L$ denote the state at time t (e.g. the position of the car). A trajectory is a series of state-action pairs, $\xi = (s_t, a_t)_{t=0}^{T}$, where T is some finite task horizon. This notation holds for one-shot tasks such as accepting or rejecting an image annotation: s_0 is the image, and a_0 is the suggested annotation.

Now, we describe an interaction session I as an MDP $:= (S_U, A_U, \mathcal{T}, \mathcal{R})$. Let A_U define the set of actions available to the user, e.g. the feedback a user can give in a particular interaction type. For example, in a reward-punishment interaction, $A_U = \{-1, +1\}$. Note that A_U and A_L need not be distinct: for demonstrations, the user may have the same action space as the learner. This is sufficient notation for discussing the curation of training sets for passive learning. For active learning, we define the state $\sigma_i \in \mathcal{S}_U$ as the parameterization of L at the ith query, as made visible to the user via means such as model weights. This is distinct from the state of the underlying learner's environment s_t, as discussed previously. The transition $\mathcal{T}$ is a property of L (e.g., gradient descent if the model is a neural network), and can be opaque to the user. The reward $\mathcal{R} : S_U \mapsto \mathbb{R}$ minimizes the difference between the desired and true output of L.

3.2 Features of the Taxonomy

Query Size (Actions). The batch size of a query is determined by the number of actions $a \in A_L$ it contains, and is given by $N \cdot T$. A query q consists of one or more trajecto-

ries ξ with finite time horizon T such that $q = \{\xi_0, .., \xi_{N-1}\}$, where N is the number of options presented to the user. In the reward-punishment interaction we have been referring to, the query presented to the user is a single ($N = 1$) example of a trajectory: $q = \{\xi_0\}$. If we were to instead use a preference interaction, the user would select one of two trajectories presented to them, such that $N = 2$ and $q = \{\xi_0, \xi_1\}$.

Response Size (Actions). The number of actions $a \in A_U$ that a user provides in response to a query q is variable in size. In the reward-punishment case, the user provides one action: either a reward, or a punishment. If we were to use a preference interaction instead, the user would provide two actions by returning a total ordering over the two options presented to them.

Intervention Options. This quantifies the user's granularity in providing feedback. In both the reward-punishment and preference interactions we have used as examples, the expectation is that the user must respond to the entirety of the query q. Therefore, the space of their intervention choices is 0; it is a coarse response. However, some interactions, such as corrections [Bajcsy *et al.*, 2018] allow the user to select subsets of $q = \{\xi_0\}$ to modify; in the lunar lander case, a user could select the entirety of ξ_0 as good and make no modifications, or adjust some chunk of $i < T$ actions. The user's intervention choice is 2^T because they have the opportunity to intervene at each time step.

Response Choice Space. This is the number of possible responses that a user can provide, given a query q, and is related to A_U. In the reward-punishment example, the user can give $|A_U| = |\{-1, +1\}| = 2$ possible responses. More generally, a user has as many response options as they have potential rewards or labels to assign. On the other hand, if we were to use a preference interaction, then $|A_U| = 2$; more generally, in ranking N options, users have $N!$ orderings to choose from.

4 User Study

We designed a mixed-design user study to find empirical differences in cognitive load and usability between interaction types. Our within-subjects independent variable, interaction type, had four levels: *Showing, Categorizing, Sorting*, and *Evaluating*. Our between-subjects independent variable, task domain, had two levels: Sequential Decision Making (henceforth SDM), and Classification.

To enable comparisons between interaction types, we selected similarly complex examples from each cluster and minimized presentation differences. We made the assumption that salient differences in user attitudes manifest even between low-complexity interactions (e.g. differences would be present between reward-punishment and 2-way preference comparisons, not just rating scales and N-way rankings). We chose the lowest-complexity, non-trivial examples of each interaction type: demonstrations with a manageable $|A_L| = 4$ for *Showing*, reward & punishment for *Categorizing*, preference comparisons for *Sorting*, and credit assignment (a subcategory of critiques) for *Evaluating* (Figure 1). We also standardized the interaction interface (e.g. the number of buttons, duration of tasks, available controls) as much as possible to minimize their impact on user attitudes.

The *SDM* task involved piloting a lunar lander to land upright between flag posts. Participants supplied or responded to a trajectory. We manually created trajectories to show to participants, and ensured an equal distribution of successful and failed trajectories. For *Showing*, participants used keyboard inputs ($|A_L| = 4$) to provide example trajectories. For *Categorizing*, participants labeled a video of a potential lunar lander trajectory with a thumbs-up or thumbs-down. For *Sorting*, participants were shown two videos of potential trajectories for a lunar lander, and chose the better one. Finally for *Evaluating*, participants were given one video of a potential lunar lander trajectory, and used a double-ended slider to select the best portion of the trajectory.

The *Classification* task consisted of a series of images to be annotated. Users provided or responded to one-word captions. We used 20 images from Pascal VOC 2012 [Everingham *et al.*, 2010] by randomly selecting one image from each of its classes. Captions to be evaluated were generated by a Keras InceptionV3 [Szegedy *et al.*, 2016] model trained on ImageNet [Deng *et al.*, 2009]. For *Showing*, participants were given an image and typed a caption of their own choosing into a textbox. For *Categorizing*, participants labeled an image-caption pair as thumbs-up or thumbs-down. For *Sorting* participants were shown two potential captions for a given image, and chose the one they felt was better. Finally, for *Evaluating*, participants were given one image-caption pair, and used a grid to select the parts of the image that best justified the proposed caption.

Hypotheses

We hypothesize that interaction types are not interchangeable with respect to their human factors:

H1 Cognitive load differs between interaction types

H2 Task completion times differ between interaction types

H3 User confidence varies between interaction types

H4 Subjective usability differs between interaction types

H5 Preferred interaction types differ between tasks

Our study is designed to identify significant differences, not to find causal relationships, but we expect that as Response Choice Space and Response Size increase, cognitive load will increase, while usability and performance suffer.

4.1 Measures

We collected metrics on cognitive load (M1, M2), performance (M3, M4), and usability (M5-M9), as well as participants' responses, and any button toggles or video replays. Participants had the opportunity to provide additional feedback, and were asked to report their age and gender.

M1 - Secondary task performance. During each interaction, participants pressed a key every time a color-changing circle turned pink (Figure 2). The longer the participants' reaction time, the greater the cognitive load [DeLeeuw and Mayer, 2008].

M2 - Paas subjective rating scale. After each interaction section, participants responded to the prompt "*How much mental effort did this interaction type demand?"* using a 9-point Likert scale [Paas, 1992] .

M3 - Primary question response time. We recorded the time between when the participant was given the stimulus (e.g. began the lunar lander game, or was first presented with an image to label) and when they submitted their response.

M4 - Self-reported confidence per query. For each query, participants responded to the prompt "*How confident are you in your answer to the primary question (not the color-changing circle) above?*" with a 4-point Likert scale. We did not include a neutral option.

M5 - Frustration. After each interaction section, participants responded to the NASA TLX [Hart and Staveland, 1988] prompt *"How insecure, discouraged, irritated, stressed, and annoyed were you?"* on a 9-point Likert scale.

M6 - Complexity. After each interaction section, participants responded to the System Usability Scale (SUS) [Brooke, 1996] prompt: *"I found this interaction type unnecessarily complex"* with a 5-point Likert scale.

M7 - Ease of Use. After each interaction section, participants responded to the SUS prompt: *"I thought this interaction type was easy to use"* with a 5-point Likert scale.

M8 - Overall Confidence. After each interaction section, participants responded to the SUS prompt: *"I felt very confident using this interaction type"* with a 5-point Likert scale.

M9 - Forced Ranking. At the study's conclusion, users answered *"My nth choice interaction type would be ..."* for their first, second, third, and fourth choice interactions.

4.2 Procedure

Participants were fully counterbalanced between all orderings of interaction types within a task domain. Participants were given instructions describing the study. They then practiced responding to the secondary task. At the beginning of each interaction type's section, participants were presented with instructions describing the interaction, and an example of a good response. Participants then practiced the interaction, including the secondary task and confidence assessment. Each interaction type's section comprised five questions presented in a sequential, but randomized, order.

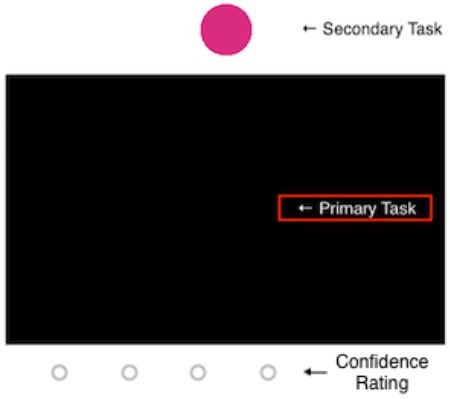


Figure 2: Participants responded to primary tasks described in Section 4.1, a secondary task (M1), and a confidence assessment (M4).

5 Results

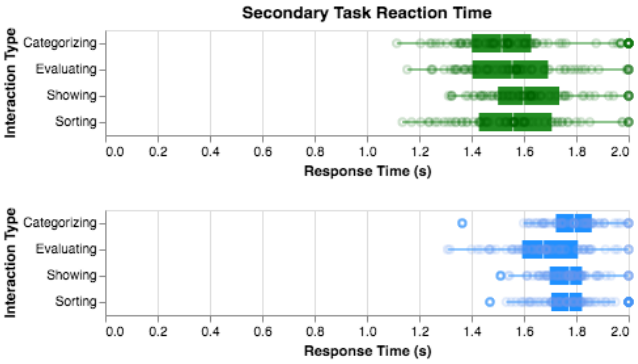
We collected data from 150 Prolific workers over the age of 18 and with approval ratings $\geq 98\%$. Partial or duplicate task completions were discarded, leaving us with 144 participants, 72 per task domain. 61.1% of the participants self-identified as male, 37.5% as female, and 1.39% as non-binary. Their ages ranged from 18 to 70 (M = 26.71, SD = 9.45). This study and recruitment procedure was approved by our Institutional Review Board. We analyzed the effects of interaction types in each domain separately, and opted not to evaluate interaction effects between domains for two reasons. First, we use Likert-type scale data which is subject to interpersonal variance and cannot be reliably compared between separate populations. Second, our goal is not to identify differences between these specific domains, but to show that task domain can affect participants' preferences.

We analyzed all ordinal data using a Friedman Test followed by a post-hoc Wilcoxon signed-rank test (Bonferonni correction $\alpha = 0.0083$). Numerical data was analyzed with a one-way repeated measures ANOVA and post-hoc pairwise Tukey analyses. We used $\alpha = 0.05$ for our analyses. Because we used only portions of NASA-TLX and SUS to avoid participant fatigue, we treat each question as an individual item.

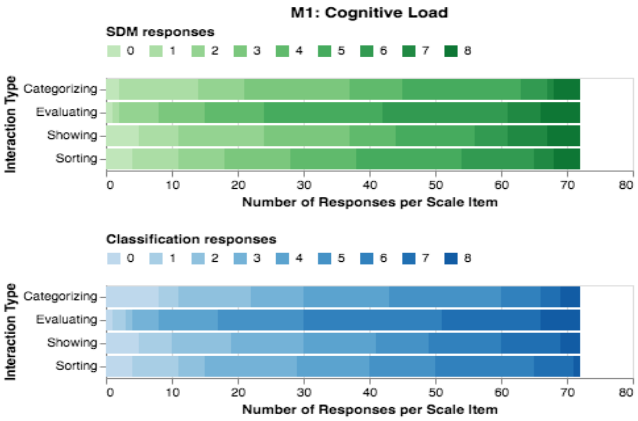
H1: Cognitive load differs between interaction types. A one-way repeated measures ANOVA revealed a statistically significant difference in secondary task reaction times (M1, Figure 3a) between interaction types ($F(3, 213) = 6.57, p < 0.001$ in SDM, and $F(3, 213) = 20.04, p < 0.001$ in Classification). In SDM, secondary reaction time was significantly longer in *Showing* as compared to *Categorizing* ($p < 0.05$). In Classification, secondary task reaction time during *Evaluating* was significantly less than in *Showing*, *Sorting* or *Categorizing* ($p < 0.01$). For completeness, we repeated this analysis after performing outlier rejection for samples more than three standard deviations from the mean: no differences were found in our results.

Differences were also found in participants' subjective assessments of the mental effort each interaction type required (M2, Figure 3b) in both domains ($\chi^2(3) = 1.3 \times 10^{-13}, p < 0.001$ in SDM, and $\chi^2(3) = 4.44 \times 10^{-15}, p < 0.001$ in Classification). In SDM, participants felt that *Sorting* was significantly harder than *Categorizing* ($p < 0.006$), and that *Evaluating* was significantly harder than *Showing*, *Sorting*, and *Categorizing* ($p < 0.001$ in all cases). In Classification, participants rated *Evaluating* as significantly harder than *Showing*, *Sorting*, and *Categorizing* ($p < 0.001$ in all cases). *The data supports H1.*

H2: Task completion times differ between interaction types. Statistically significant differences were found in response times (M3, Figure 4) between interaction types in both domains ($F(3, 213) = 28.79, p < 0.001$ in SDM, $F(3, 213) = 166.29, p < 0.001$ in Classification). In SDM,



(a) Objective cognitive load; higher values indicate greater load.



(b) Subjective cognitive load; darker values indicate greater load.

Figure 3: Cognitive load (H1) was measured both objectively (secondary task reaction time) and subjectively.

response times were significantly greater in *Sorting* as compared to *Showing* and *Categorizing*, and in *Evaluating* as compared to any other interaction type ($p < 0.01$ for both). In Classification, participants' response times to *Evaluating* were significantly greater than to any other interaction type ($p < 0.01$ in all cases). Figure 4 demonstrates these findings visually. When we repeated this analysis after performing outlier rejection for samples more than three standard deviations from the mean, our results largely stayed the same: we additionally found that *Showing* interactions in the Classification domain took significantly longer than *Categorizing* ($\alpha < 0.05$). *The data supports H2.*

H3: User confidence varies between interaction types. Median per-trial confidence scores (M4, Figure 5) were significantly different between interaction types in both domains ($\chi^2(3) = 49.24, p < 0.001$ in SDM, and $\chi^2(3) = 102.91, p < 0.001$ in Classification). In SDM, participants were significantly more confident in their responses to *Showing* and *Categorizing* than *Sorting* or *Evaluating* ($p < 0.01$ in all cases). In Classification, participants were more confident in their responses to *Showing* than to any other interaction ($p < 0.01$ in all cases). They were also more confident in their responses to *Categorizing* than to *Sorting* and *Evaluating* ($p < 0.01$ in both cases). *The data supports H3.*

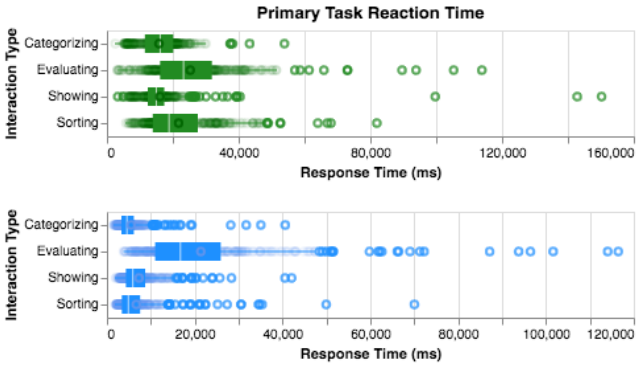


Figure 4: Time taken to complete the primary interaction task (H2).

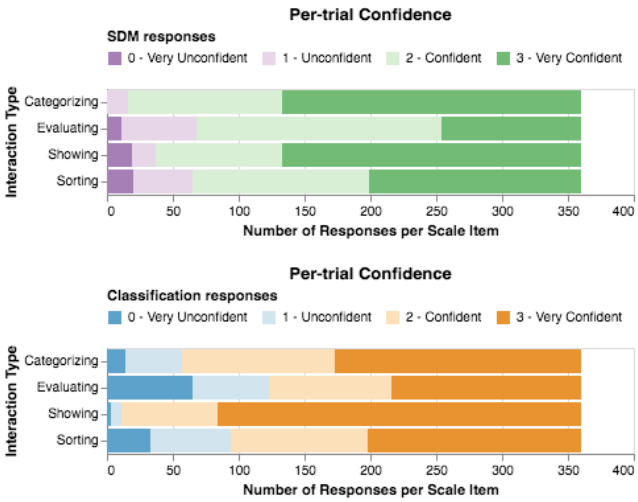


Figure 5: Per-trial confidence in response quality (H3).

H4: Subjective usability differs between interaction types. Significant differences were found in participants' ratings of frustration ($\chi^2(3) = 27.07, p < 0.001$ in SDM, $\chi^2(3) = 50.94, p < 0.001$ in Classification), perceptions of complexity ($\chi^2(3) = 41.68, p < 0.001$ in SDM, and $\chi^2(3) = 69.30, p < 0.001$ in Classification), ease of use ($\chi^2(3) = 33.14, p < 0.001$ in SDM, and $\chi^2(3) = 54.19, p < 0.001$ in Classification), and confidence with the interaction type ($\chi^2(3) = 28.66, p < 0.001$ in SDM, and $\chi^2(3) = 55.63, p < 0.001$ in Classification). These results, corresponding to M5 through M8, are shown in Figures 6 through 9.

Participants felt more frustrated by *Evaluating* than any other interaction in both task domains ($p < 0.00145$ in all cases). They perceived *Evaluating* as more unnecessarily complex than any other interaction in both task domains as well ($p < 0.001$ in all cases); in Classification, they also perceived *Sorting* as unnecessarily more complex than *Categorizing* ($p < 0.00785$). Correspondingly, participants found *Evaluating* to be less easy to use than any other interaction type in both task domains ($p < 0.001$ in all cases). In SDM, participants were more confident with *Showing*, *Sorting*, and *Categorizing* over *Evaluating* ($p < 0.001$ in all cases). In Classification, participants felt more confident using *Show-*

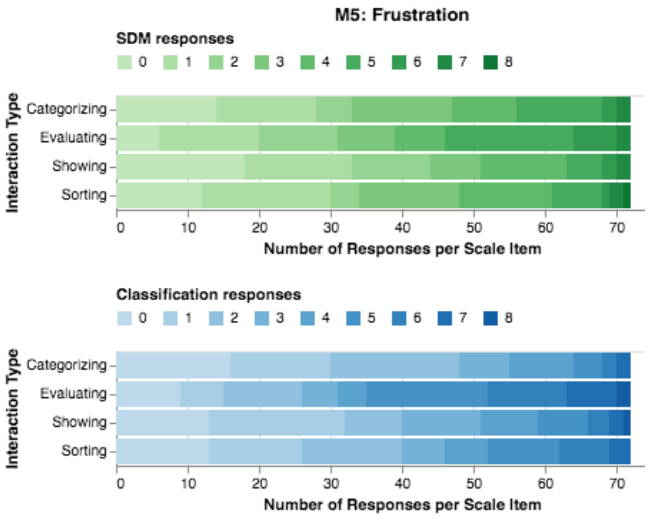


Figure 6: Responses to subjective measures of frustration (H4). Darker colors denote greater frustration.

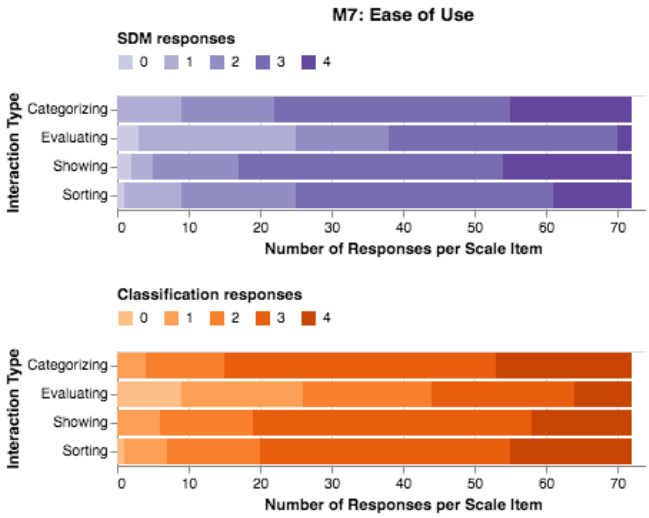


Figure 8: Responses to subjective measures of ease of use (H4). Darker colors denote greater ease of use.

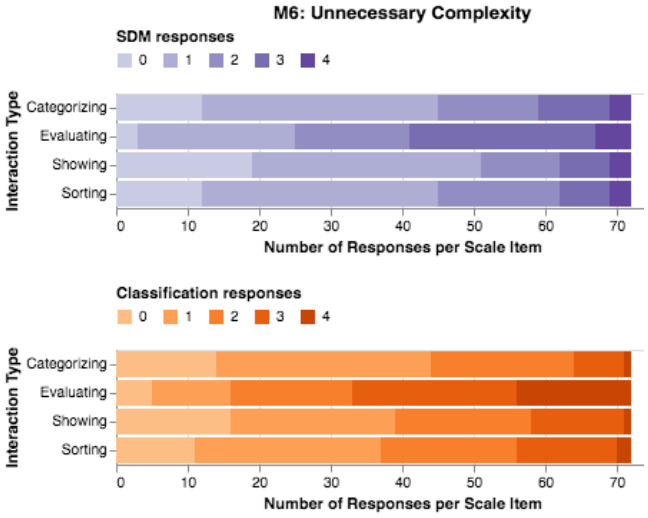


Figure 7: Responses to subjective measures of complexity (H4). Darker colors denote higher perceived complexity.

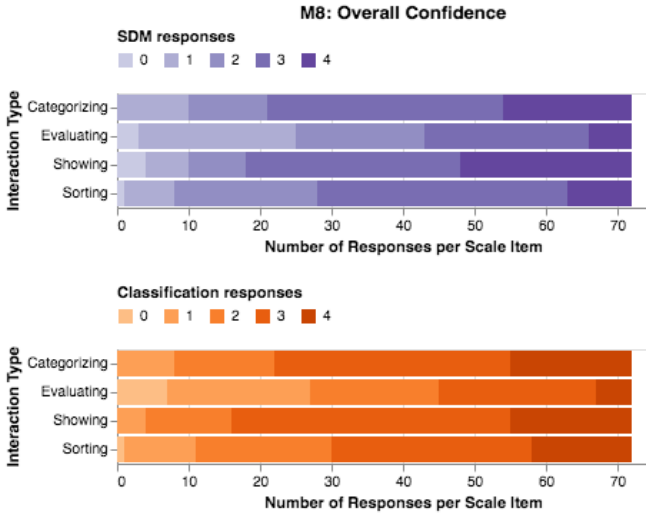


Figure 9: Responses to subjective measures of confidence (H4). Darker colors denote greater perceived confidence.

ing than *Sorting* ($p < 0.00230$) or *Evaluating* ($p < 0.001$). They also felt more confident using either *Sorting* or *Categorizing* over *Evaluating* ($p < 0.001$ in both cases). *The data supports H4.*

H5: Preferred interaction types differ between tasks. We tallied participants' preferred interaction types (M9) using a Condorcet method. We found that in SDM, participants preferred *Showing*, *Sorting*, *Categorizing*, and then *Evaluating*. In Classification, they preferred *Categorizing*, *Sorting*, *Showing*, and then *Evaluating*. *The data supports H5.*

6 Discussion

Our results show that interaction types are differently cognitively loading and usable, and may variably impact performance as estimated via task completion times and self-assessed confidence. Participants rated *Evaluating* interactions as requiring the most cognitive effort, being the most frustrating, the most unnecessarily complex, least easy to use, and inspiring the least confidence. Objectively, they also took the longest time to complete *Evaluating* tasks. In SDM, *Sorting* took longer than *Showing* and *Categorizing*. In Classification, participants felt *Sorting* was more unnecessarily complex than *Categorizing* and were less confident using it than *Showing*. This suggests that *Categorizing* is preferable to *Sorting*, which is preferable to *Evaluating*. This corresponds to our expectation that as an interaction's Response Choice Space and Response Size increases, its usability decreases.

Unexpectedly, *Showing*, has the largest Response Choice Space and was among the easiest to use. This may be due

to cognitive shortcuts: when guiding the lunar lander, users may not be processing all possible trajectories, nor are they thinking of every word they know in order to caption images. Thus, our big-$\mathcal{O}$ estimates may have been too coarse to capture the nuances of a user's *perceived* response space.

We also found a disagreement between participants' subjective assessment of mental effort and their objective secondary task performance, as in prior work [DeLeeuw and Mayer, 2008]. This could indicate that there are additional, unknown factors that affect perceived mental effort. We did also observe a relationship between participants' primary task reaction times and subjective assessments of cognitive load, indicating that they took longer on cognitively loading tasks. This may have given them more opportunities to respond to the secondary task, influencing their reaction times.

Pre-existing notions that interaction types such as *Evaluating* and *Sorting* might be more user-friendly than others (particularly *Showing*), because they require fewer inputs from a user, were not supported in the two domains we evaluated. Furthermore, differences existed in participants' preferred interactions between the task domains, despite our standardization of interaction types within and between them. Future work is required to understand how properties of a task domain influence interactions.

This work is one step towards developing a principled understanding of the algorithmic and human-factors components of learning interactions. In particular, it is a necessary step towards understanding the trade-off between the expected informativeness of a learning interaction, and a user's ability to provide high quality feedback. As data-gathering needs increase in scale and across domains, understanding this relationship will expand our ability to design learning interactions that not only accommodate the needs of learning agents, but also leverage the capabilities of human teachers.

Acknowledgements

We thank our anonymous reviewers, as well as Aditya Dhawale, Tesca Fitzgerald, Misha Khodak, Ada Taylor and the HARP Lab at CMU for their feedback and advice. We also thank Meghna Behari for helping with coding the user study during her internship. This work is supported in part by the Office of Naval Research (N00014-18-1-2503). The Classification image in Figure 1 is "Inside Whitehaven library" by librariesteam and is licensed with CC BY 2.0 [Creative Commons, 2004].

References

[Abbeel and Ng, 2004] Pieter Abbeel and Andrew Y Ng. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the twenty-first international conference on Machine learning*, page 1. ACM, 2004.

[Amershi *et al.*, 2014] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. Power to the people: The role of humans in interactive machine learning. *Ai Magazine*, 35(4):105–120, 2014.

[Bain and Sammut, 1995] Michael Bain and Claude Sammut. A framework for behavioural cloning. In *Machine Intelligence 15*, pages 103–129, 1995.

[Bajcsy *et al.*, 2018] Andrea Bajcsy, Dylan P Losey, Marcia K O'Malley, and Anca D Dragan. Learning from physical human corrections, one feature at a time. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 141–149. ACM, 2018.

[Bıyık *et al.*, 2020] Erdem Bıyık, Malayandi Palan, Nicholas C Landolfi, Dylan P Losey, Dorsa Sadigh, et al. Asking easy questions: A user-friendly approach to active reward learning. In *Conference on Robot Learning*, pages 1177–1190, 2020.

[Brooke, 1996] John Brooke. Sus: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, and A. L. McClelland, editors, *Usability Evaluation in Industry*. London: Taylor and Francis, 1996.

[Bullard *et al.*, 2018] Kalesha Bullard, Andrea L Thomaz, and Sonia Chernova. Towards intelligent arbitration of diverse active learning queries. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6049–6056. IEEE, 2018.

[Cakmak and Thomaz, 2012] Maya Cakmak and Andrea L Thomaz. Designing robot learners that ask good questions. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 17–24. ACM, 2012.

[Chandler and Sweller, 1991] Paul Chandler and John Sweller. Cognitive load theory and the format of instruction. *Cognition and instruction*, 8(4):293–332, 1991.

[Creative Commons, 2004] Creative Commons. Attribution 2.0 generic (cc by 2.0), 2004. [Online; accessed 11-May-2021].

[Cui and Niekum, 2018] Yuchen Cui and Scott Niekum. Active reward learning from critiques. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6907–6914. IEEE, 2018.

[Daniel *et al.*, 2015] Christian Daniel, Oliver Kroemer, Malte Viering, Jan Metz, and Jan Peters. Active reward learning with a novel acquisition function. *Autonomous Robots*, 39(3):389–405, 2015.

[DeLeeuw and Mayer, 2008] Krista E DeLeeuw and Richard E Mayer. A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of educational psychology*, 100(1):223, 2008.

[Deng *et al.*, 2009] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[Everingham *et al.*, 2010] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[Fürnkranz and Hüllermeier, 2011] Johannes Fürnkranz and Eyke Hüllermeier. *Preference Learning and Ranking by Pairwise Comparison*, pages 65–82. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[Hadfield-Menell *et al.*, 2016] Dylan Hadfield-Menell, Anca Dragan, Pieter Abbeel, and Stuart Russell. The off-switch game. *arXiv preprint arXiv:1611.08219*, 2016.

[Hart and Staveland, 1988] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[Jain *et al.*, 2015] Ashesh Jain, Shikhar Sharma, Thorsten Joachims, and Ashutosh Saxena. Learning preferences for manipulation tasks from online coactive feedback. *The International Journal of Robotics Research*, 34(10):1296–1313, 2015.

[Jeon *et al.*, 2020] Hong Jun Jeon, Smitha Milli, and Anca D Dragan. Reward-rational (implicit) choice: A unifying formalism for reward learning. *arXiv preprint arXiv:2002.04833*, 2020.

[Kittur *et al.*, 2013] Aniket Kittur, Jeffrey V Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1301–1318, 2013.

[Kober *et al.*, 2013] Jens Kober, J Andrew Bagnell, and Jan Peters. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11):1238–1274, 2013.

[Lease, 2011] Matthew Lease. On quality control and machine learning in crowdsourcing. *Human Computation*, 11(11), 2011.

[Lin *et al.*, 2014] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[Longo, 2018] Luca Longo. Experienced mental workload, perception of usability, their interaction and impact on task performance. *PloS one*, 13(8):e0199661, 2018.

[Ng *et al.*, 2000] Andrew Y Ng, Stuart J Russell, et al. Algorithms for inverse reinforcement learning. In *ICML*, volume 1, page 2, 2000.

[O'Connor and Cheema, 2018] Kieran O'Connor and Amar Cheema. Do evaluations rise with experience? *Psychological Science*, 29(5):779–790, 2018.

[Paas, 1992] Fred G Paas. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of educational psychology*, 84(4):429, 1992.

[Palan *et al.*, 2019] Malayandi Palan, Nicholas C Landolfi, Gleb Shevchuk, and Dorsa Sadigh. Learning reward functions by integrating human demonstrations and preferences. In *Robotics: Science and Systems (RSS)*, 2019.

[Prewett *et al.*, 2010] Matthew S Prewett, Ryan C Johnson, Kristin N Saboe, Linda R Elliott, and Michael D Coovert. Managing workload in human–robot interaction: A review of empirical studies. *Computers in Human Behavior*, 26(5):840–856, 2010.

[Ramachandran and Amir, 2007] Deepak Ramachandran and Eyal Amir. Bayesian inverse reinforcement learning. In *IJCAI*, volume 7, pages 2586–2591, 2007.

[Real *et al.*, 2017] Esteban Real, Jonathon Shlens, Stefano Mazzocchi, Xin Pan, and Vincent Vanhoucke. Youtube-boundingboxes: A large high-precision human-annotated data set for object detection in video. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5296–5305, 2017.

[Sadigh *et al.*, 2017] Dorsa Sadigh, Anca D Dragan, Shankar Sastry, and Sanjit A Seshia. Active preference-based learning of reward functions. In *Robotics: Science and Systems (RSS)*, 2017.

[Schaal, 1999] Stefan Schaal. Is imitation learning the route to humanoid robots? *Trends in cognitive sciences*, 3(6):233–242, 1999.

[Sweller, 1988] John Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285, 1988.

[Szegedy *et al.*, 2016] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[Wirth *et al.*, 2017] Christian Wirth, Riad Akrour, Gerhard Neumann, and Johannes Fürnkranz. A survey of preference-based reinforcement learning methods. *The Journal of Machine Learning Research*, 18(1):4945–4990, 2017.

[Ziebart *et al.*, 2008] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, and Anind K Dey. Maximum entropy inverse reinforcement learning. In *AAAI*, volume 8, pages 1433–1438. Chicago, IL, USA, 2008.