Learning versus performance in social contexts

Erica J. Yoon*, **Kyle MacDonald***, **Mika Asaba**, **Hyowon Gweon**, and **Michael C. Frank**
{ejyoon, kylem4, masaba, hyo, mcfrank} @stanford.edu
Department of Psychology, Stanford University
*These authors contributed equally to this work.

Abstract

...

Keywords: Learning; social context; information gain; OED; self-presentation; goal tradeoff

Introduction

Imagine that you are a novice cook and you have to decide what meal to prepare for a first date. Should you choose a recipe that you have tried before or should you attempt to make something new? While the familiar recipe has a high chance of ensuring a good meal, you are less likely to discover a new, delicious dish. The new recipe might taste even better, but it has a higher chance of turning out awful since you have never made it before. Now, consider how your decision making might change if instead of making the meal for first date, you were preparing it for a charitable cooking teacher whom you trust.

Scenarios like this, capture an "explore-exploit" dilemma (Sutton & Barto, 1998), in which we have to choose between actions that could (a) lead to an overt, readily accessible reward based on what we already know (*exploitation*) or (b) result in the discovery of new information (*exploration*). This decision of whether to explore or exploit is directly related to the relative strength of our goals within a particular context. In the cooking example, should I prioritize the goal of learning by cooking the new recipe, or should I emphasize the performance goal by preparing the tried and true meal? The key insight motivating the work reported here is that features of the social context play a fundamental role in the goals we consider and the relative weighting of each goal. And, in turn, the goals that we consider shape the actions take. We present a formal account to integrate social reasoning processes with decision making in the context of learning-performance tradeoff, as a case study of the explore-exploit dilemma.

We situate our integrative account within two theoretical frameworks: *active learning* and *pragmatic social reasoning*. Active learning refers to situations where people are given control over the sequence of information in a learning context (e.g., verbal question asking to elicit informative responses). The key assumption is that learners will maximize the usefulness of their actions by gathering information that is especially helpful for their own learning. The effects of active learning have been the focus of much empirical work in education (Grabinger & Dunlap, 1995; Prince, 2004), machine learning (Ramirez-Loaiza, Sharma, Kumar, & Bilgic, 2017; Settles, 2012), and cognitive psychology (Castro et al., 2009; Chi, 2009), with the common finding that active contexts lead to more rapid learning when compared to passive contexts where people do not have control over the flow of information.

Work on exploratory actions in active learning often isolates people's information goals by removing the learner from any kind of social context. In contrast, real-world learning is characterized by contexts where there are teachers, peer learners, or other individuals who can directly influence the utility of information gathering actions. Consider that there is now a large body of evidence suggesting that social reasoning processes can change the computations (i.e., inferences) that support learning from evidence. For example, children learn at different rates after observing the same evidence depending on whether they thought the behavior was accidental (less informative) or intentional (more informative). Moreover, adults and children will make even stronger inferences if they believe that another person selected their actions with the goal of helping them learn (i.e., teaching) (Shafto, Goodman, & Frank, 2012).

But how can we begin to understand the role of *social* factors in self-directed learning contexts? Answering this question represents a significant step because explore-exploit decisions often unfold within fundamentally social contexts where people must integrate the value of social goals and information goals when deciding what to do next. Consider that actions that maximize learning are inherently risky – leading to mistakes and struggles – which could be difficult to undertake in with someone else present. Thus, a learner might prioritize actions that maintain their reputation: looking more attractive, knowledgeable, or competent to the observer. In this case, the learner is leveraging psychological reasoning processes to infer what the observer thinks based on their action (cook a meal) and its outcome (a delicious meal). If the learner is worried about the observer's beliefs, then the they might choose actions that maximize the probability of maintaining a favorable impression in the eyes of the observer ('That meal was delicious, so he must be a good cook!'). Indeed, when the value of self-presentation is highlighted, children avoid opportunities that will help them learn new things but pose risks for public mistakes (Dweck & Leggett, 1988; Elliott & Dweck, 1988).

We present the first computational model of the learning-performance tradeoff in a self-directed learning context with social factors. We model a learner who considers his learning goal versus performance goal, which may be influenced by the social context (e.g. the presence of another individual whom he wants to impress). We then look at adult par-

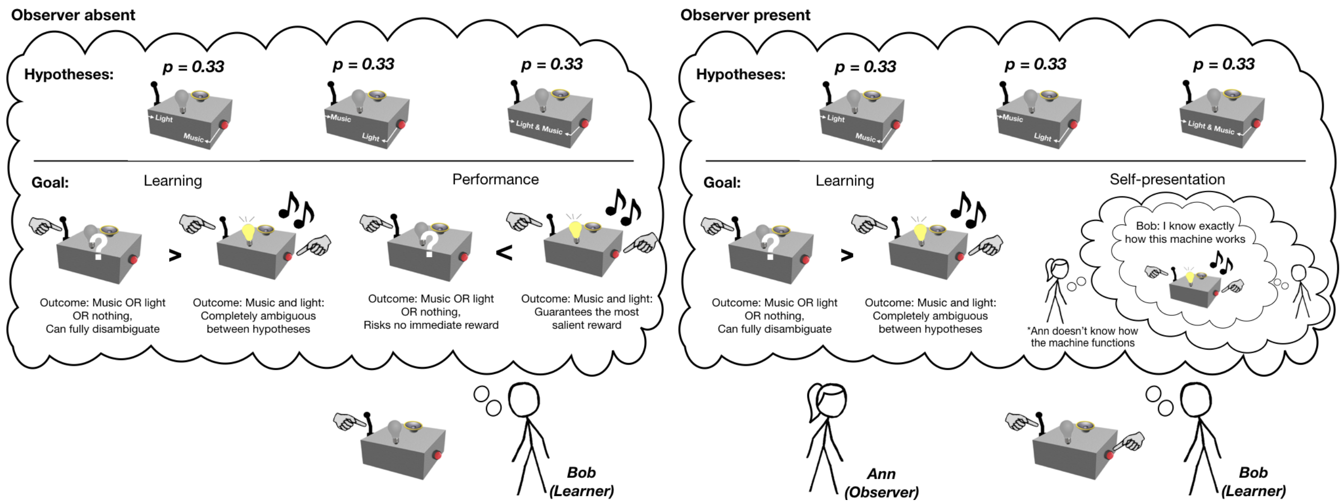

Figure 1: Diagram of the computational model: The learner considers possible hypotheses and his contextual goals. When an observer is absent, he considers his learning goal (to maximize information gain) and performance goal (e.g. to play music) and decides on an action. When an observer is present, his decision for an action is based on his learning goal vs. presentational goal (to have the observer infer his competence).

ticipants' action choices within a minimalistic self-directed learning task representing different goals and social contexts, and show that people's choices are consistent with predictions of our model.

Computational model

To start to examine people's exploration-exploitation tradeoff, we situate the model and paradigm in a simplistic learning environment. The learner in our model is to act on a toy, and can choose between two kinds of actions that will each lead to one outcome (new discovery) or the other (overt reward). The learner's action rests on his goals to explore versus exploit, and is determined in part by the presence or absence of another person he cares about (i.e. his boss)[1].

A key assumption underlying inferences in recent Bayesian models of human social cognition is that people act approximately optimally given a utility function (e.g. Goodman & Frank, 2016; Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016). Our model adopts the same utility-theoretic approach, and assumes an approximately optimal agent, who reasons about the utility function that represents a combination of multiple goals. In a recent model of polite language production (Yoon, Tessler, Goodman, & Frank, 2017), the utility function comprised a weighted combination of multiple utilities (goals) considered by the speaker, reflecting a principled tradeoff between different communicative goals (e.g. to be informative, to be kind, and to appear to be a helpful speaker). We use a similarly structured utility function that reflects different goals that a learner has in a social learning context. Specifically, we model how a person may make a decision to act based on his desire to learn how a toy works (*learning utility*), to make the toy operate and perform a given function (*performance utility*), or to present himself as a competent individual who knows how to make the toy work (*presentational utility*; see the model diagram in Figure 1).

First, the *learning utility* symbolizes the goal to learn new information, which in our paradigm specifically is associated with figuring out how a given toy works. The learning utility is formally represented by an OED model (Lindley, 1956; "Optimal Experiment Design"; Nelson, 2005), which quantifies the *expected utility* of different information seeking actions. Here we follow the mathematical details of the OED approach as outlined in Coenen, Nelson, & Gureckis (2017) that was implemented in our model. The set of queries, each realized through taking an action, is defined as $Q_1, Q_2, ..., Q_n = Q$. The expected utility of each query ($EU(Q)$) is a function of two factors: (1) the probability of obtaining a specific answer $P(a)$ weighted by (2) the usefulness of that answer for achieving the learning goal $U(a)$.

$$EU(Q) = \sum_{a \in q} P(a)U(a)$$

There are a variety of ways to define the usefulness function to score each answer (for a detailed analysis of different approaches, see Nelson (2005)). One standard method is to use *information gain*, which is defined as the change in the learner's overall uncertainty (difference in entropy) before and after receiving an answer. This information gain is then the usefulness of the answer to the query, and thus is equal to the learning utility:

$$U_{learning} = U(a) = ent(H) - ent(H|a)$$

[1] From here on, we use a male pronoun for Bob, the learner, and female pronoun for Ann, the boss and observer.

where $ent(H)$ is defined using Shannon entropy[2]. MacKay (2003), which provides a measure of the overall amount of uncertainty in the learner's beliefs about the candidate hypotheses.

$$ent(H) = -\sum_{a \in A} P(h)log_2 P(h)$$

The conditional entropy computation is the same, but takes into account the change in the learner's beliefs after seeing an answer.

$$ent(H|a) = -\sum_{h \in H} P(h|a)logP(h|a)$$

To calculate the change in the learner's belief in a hypothesis $P(h|a)$, we use Bayes rule.

$$P(h|a) = \frac{P(h)P(a|h)}{P(a)}$$

The learner performs the expected utility computation for each query in the set of possible queries and picks the one that maximizes utility. In practice, the learner considers each possible answer, scores the answer with the usefulness function, and weights the score using the probability of getting that answer. In our paradigm, a learner thinking about the learning utility considers acting on the toy one way over another, and computes how informative a given answer should be in reducing uncertainty about how the toy works.

Second, the *performance utility* is the utility of successfully making the toy operate. Specifically within our current paradigm, the performance utility is the expected utility of music playing (*m*) given the learner's action *a*.

$$U_{performance} = P_L(m|a)$$

Thus, performance utility is maximized by taking an action that is most likely to make the toy "go" and play music, which is the operation of interest.

When there is no observer present, the learner considers the tradeoff between the learning utility and performance utility, and he determines his action based on a weighted combination of the two utilities:

$$U(a;\phi;obs = no) = \phi \cdot U_{learning} + (1-\phi) \cdot U_{performance},$$

where ϕ is a mixture parameter governing the extent to which the learner prioritizes information gain over making the toy play music.

[2]Shannon entropy is a measure of unpredictability or amount of uncertainty in the learner's probability distribution over hypotheses. Intuitively, higher entropy distributions are more uncertain and harder to predict. For example, if the learner believes that all hypotheses are equally likely, then they are in a state of high uncertainty/entropy. In contrast, if the learner firmly believes in one hypothesis, then uncertainty/entropy is low.

When there is another person present to observe the learner's action, this observer *O* reasons about the competence *c* of the learner *L* which is equal to whether the learner was able to make the toy work.

$$P_O(c) \propto P_L(m|a)$$

The learner thinks about how the observer infers the learner's competence, and his *presentational* utility is based on maximizing the apparent competence inferred by the observer.

$$U_{presentation} = P_O(c)$$

Thus, when there is an observer present, the learner considers the tradeoff between the learning utility and presentational utility:

$$U(m;a;\phi;obs = yes) = \phi \cdot U_{learning} + (1-\phi) \cdot U_{presentational}$$

Based on the utility functions above, the learner (*L*) chooses his action *a* approximately optimally (as per optimality parameter λ) given his goal weight and observer presence.

$$P_L(a|\phi, obs) \propto \exp(\lambda \cdot \mathbb{E}[U(a;\phi;obs)])$$

Experiment 1

In our experimental paradigms, we wanted to look at people's action choices in a minimalistic self-directed learning environment with varying goals and social contexts. The task would involve choosing between an action that is informative but potentially not immediately rewarding, and an action that is certainly immediately rewarding but not informative at all.

In Experiment 1, we first wanted to confirm that participants would choose different actions depending on goals that we highlighted. We were also interested to see what people would choose to do when no goal is specified and they are free to explore. Importantly, participants were limited to making only a single action, which meant the opportunity cost for the alternative action was at its highest. Thus, participants were asked to imagine that they needed to act on a toy with an uncertain causal mechanism, and we assigned participants to different goal conditions: learning goal (i.e. learn how the toy works); performance goal (e.g. make the toy play music); presentation goal (i.e. impress their boss); and no goal.

We hypothesized that participants would choose an informative action more often in the following order of goal conditions (decreasing): learning, no goal, performance, and presentation[3].

Method

Participants We recruited 196 participants (roughly 50 for each condition) on Amazon's Mechanical Turk. To participate participants were required to have an IP addresses in the

[3]Our hypothesis and method were pre-registered prior to data collection on the Open Science Framework (https://osf.io/kcjau)

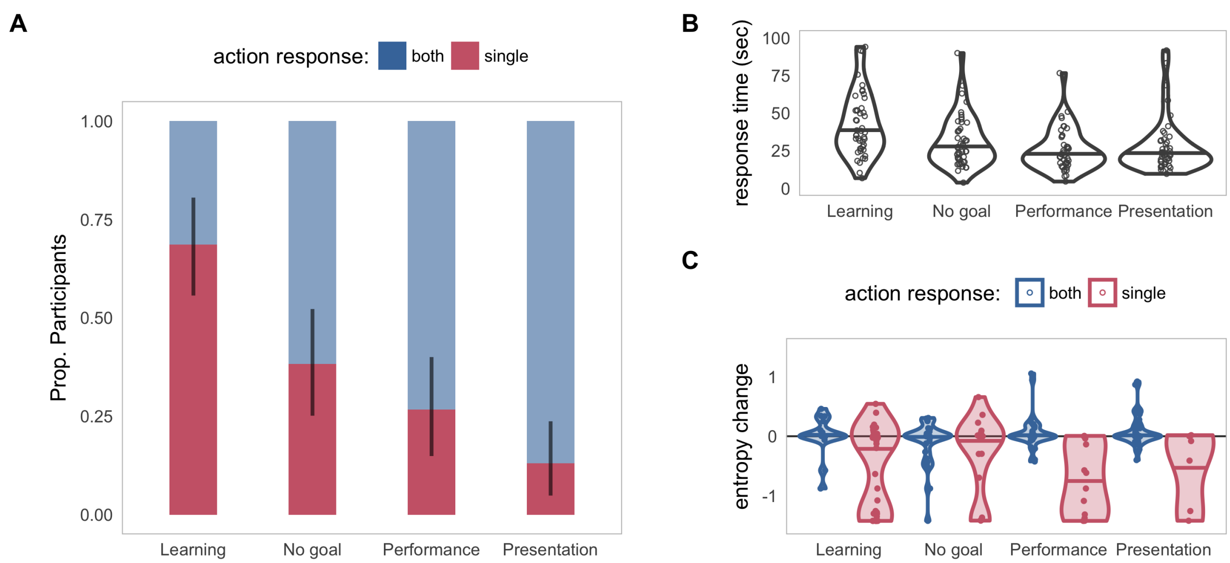

Figure 2: Behavioral results for E1. Panel A shows the proportion of action decisions for each goal condition. Error bars represent 95% binomial confidence intervals computed using Bayesian inference. Panel B shows violin plots of participants' response times on the action decisions. Each point represents a participant with the width of the violin representing the density of the data at that value. Panel C shows violin plots of participants' belief change (entropy) as a function of condition. Lower values represent higher certainty after selecting an action. Color in panels A and C represent the type of action participants selected.

United States and a task approval rate above 85%. We excluded 7 participants who failed to answer at least two out of three manipulation check questions correctly (see Procedure section for details on the manipulation check), and thus the remaining 189 participants were included in our final analysis.

Stimuli and Design We presented images of three different toys that look very similar but each work in different ways, and provided instructions for them (see top of Fig. 1 for what the toys looked like).

- Toy 1 instructions were: *"Pull the handle on the left to turn on the light. Press the button on the right to play music. Doing both produces both effects at the same time."*

- Toy 2 instuctions were: *"Pull the handle on the left to play music. Press the button on the right to turn on the light. Doing both produces both effects at the same time."*

- Toy 3 instuctions were: *"Pull the handle on the left AND press the button on the right to turn on the light and play music at the same time. The button press or handle pull on its own doesn't produce any effect."*

Thus, doing both button press and handle pull was immediately rewarding but uninformative (as it does not disambiguate the causal mechanism in any way), whereas either of the single actions was completely disambiguating, but was uncertain to produce an immediate outcome. Each toy had a label at the front, indicating which action(s) will make the toy operate, and with which outcome effect.

We asked participants to act on one of these toys; importantly, the given toy was missing its label, such that partcipants could not know whether the toy was Toy 1, 2 or 3. We assigned participants into four goal conditions. For participants in *learning*, *performance*, and *presentation* conditions,we asked them to imagine that they were children's toy developers and that one day their boss approached them. We then instructed participants to: figure out the correct label for the toy (*learning* condition); make the toy play music (or turn the light on; *performance* condition); or impress their boss and show that they are competent (*presentation* condition). In *no-goal* condition, participants were asked to select an action. We asked participants to select an action they would like to try out on the toy in order to accomplish the specified goal, out of three possible actions: to "press the button", "pull the handle", or "press the button and pull the handle." We randomly assigned each participant to one of the three goal conditions, and randomized the order of actions to choose from.

Procedure We first introduced participants to the task, and showed them a picture of a possible toy with labels on its different parts. Then they read instructions for each of the three toy types. We presented Toy 1 and Toy 2 instructions in a randomized order first, and then Toy 3 instructions. Afterwards, they were asked what they would do to make the toy operate as manipulation check (e.g. "How would you make the toy play music?"). We asked participants to rate prior likelihood that an unknown toy is Toy 1, 2, or 3, to use as priors for our model. Participants then read a scenario for one of the three goal conditions, followed by the question: "If you only had one chance to try a SINGLE action to [pursue the specified goal], which action would you want to take? You will get a 10 cent bonus after submitting the HIT if you [achieve the given goal]." After selecting one of three possible actions to perform on the toy and seeing that the toy successfully played music, participants were asked again to rate the likelihood that the unlabeled toy was each of the three possible toys.

Results and discussion

Analysis plan First, we present behavioral analyses of participants' (1) action decisions, (2) action decision times, and (3) belief change (i.e., learning).[4] Decision times correspond to the latency to make an action selection as measured from the start of the action decision trial (all RTs were analyzed in log space). We quantified participants' beliefs about the possible toy designs using entropy, and belief change was measured as the difference in entropy before and after selecting an action.

We used the `rstanarm` (Gabry & Goodrich, 2016) package to fit Bayesian regression models estimating the differences across conditions. We report the uncertainty in our point estimates using 95% Highest Density Intervals (HDI). The HDI provides a range of credible values given the data and model. All analysis code for the statistical models can be found in the online repository for this project: `https://github.com/kemacdonald/soc-info/R/03_models.Rmd`.

Next, we present a Bayesian Data Analysis of the correspondence between our computational model and participants action responses. [FIXME – say more about the BDA?]

Action decisions: We modeled action decisions using a logistic regression specified as $action \sim goal\_condition$ with the No-Goal condition as the reference category. Participants' tendency to select a "single" action varied across conditions in the predicted pattern (see Panel A of Fig 2), with the highest proportion occuring in the Learning context ($M_{learn} = 32\%$, [20%, 45%]), followed by the No Goal context ($M_{noGoal} = 62\%$, [48%, 74%]), then Performance ($M_{perform} = 73\%$, [59%, 85%]), and the fewest single actions in the Presentation condtion ($M_{present} = 86\%$, [76%, 95%]).

Compared the No-Goal condition, participants selected the single action at a greater rate in the Learning context (β = -1.27, [-2.08, -0.5]) and at lower rate in the Presentation context (β = 1.43, [0.44, 2.52]), with the null value of zero difference condition falling well outside the 95% HDI. In contrast, participants in the Performance condition selected actions at at similar rate (β = 0.53, [-0.33, 1.46]) with the 95% HDI including the null value.

Action decision times: We analyzed response times in log space using the same model specification. Panel A of Figure 2 shows the full RT data distribution. On average, participants took 27.82 seconds to generate a response in the No-goal condidition. Participants took on on average 1.48 seconds longer to generate a decision in the Learning condition (β = 0.39, [1.15, 1.88]) but produced similar response times in the Performance (β = -0.15, [0.66, 1.11]) and Presentation (β = -0.11, [1.15, 1.14]) conditions.

Belief change

BDA model-data fits

Experiment 2

In Experiment 1, we saw that participants made different action choices depending on the goal conditions, as we previously predicted.In Experiment 2, we manipulated goals as well as social contexts, fully crossing the different goal conditions with the presence/absence of the boss, to see whether the social context affects people's decision making differently in each goal condition.

We hypothesized that social pressure should increase presentation-oriented, immediately-rewarding actions in the learning and no-goal conditions, but not in the performance condition in which they are already specified a goal in the same direction.

Method

Participants We recruited 347 participants (roughly 50 for each condition) on Amazon's Mechanical Turk. To participate participants were required to have an IP addresses in the United States and a task approval rate above 85%. We excluded -129 participants who failed to answer at least two out of three manipulation check questions correctly, and thus the remaining 325 participants were included in our final analysis.

Stimuli and Design The stimuli and design were identical to Experiment 1, except we had seven different goal $\times$ social conditions. Goals remained identical to ones presented in Experiment 1; social conditions varied depending on whether the boss was present in the story (*social*) or she was absent (*non-social*). Thus, the conditions from Experiment 1 were used as *social-learning*, *social-performance*, *social-presentation*, and *non-social-no-goal* conditions in Experiment 2. We added three more conditions: *non-social-learning*, *non-social-performance*, and *social-no-goal*. Note that we did not have *non-social-presentation* condition, because presentation goal by definition was to present oneself as competent to and impress another person.

Procedure The procedure was identical to Experiment 1.

Results and discussion

Action selections

Response times

Entropy change

BDA model-data fits

[4] See TODO for a pre-registration of the analysis plan.

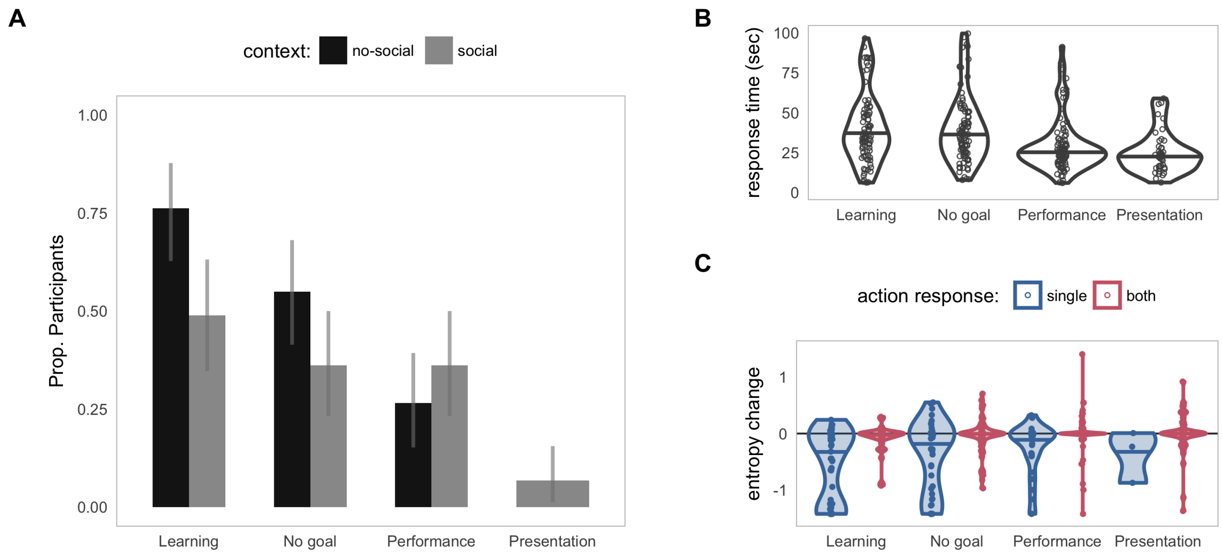

Figure 3: Behavioral results for E2. Panel A shows actions decisions with color representing social context. Panel B shows decision times. Panel C shows belief change. All other plotting conventions are the same as Figure 2.

General Discussion

Acknowledgements

This work was supported by NSERC postgraduate doctoral scholarship PGSD3-454094-2014 to EJY and an NSF GRFP to KM [FIXME].

References

Castro, R. M., Kalish, C., Nowak, R., Qian, R., Rogers, T., & Zhu, X. (2009). Human active learning. In *Advances in neural information processing systems* (pp. 241–248).

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105.

Coenen, A., Nelson, J. D., & Gureckis, T. M. (2017). Asking the right questions about human inquiry.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*(2), 256.

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, *54*(1), 5.

Gabry, J., & Goodrich, B. (2016). Rstanarm: Bayesian applied regression modeling via stan. r package version 2.10. 0.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829.

Grabinger, R. S., & Dunlap, J. C. (1995). Rich environments for active learning: A definition. *ALT-J*, *3*(2), 5–34.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The nave utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*(8), 589–604.

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 986–1005.

MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.

Nelson, J. D. (2005). Finding useful questions: On bayesian diagnosticity, probability, impact, and information gain. *Psychological Review*, *112*(4).

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, *93*(3), 223–231.

Ramirez-Loaiza, M. E., Sharma, M., Kumar, G., & Bilgic, M. (2017). Active learning: An empirical study of common baselines. *Data Mining and Knowledge Discovery*, *31*(2), 287–313.

Settles, B. (2012). Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, *6*(1), 1–114.

Shafto, P., Goodman, N. D., & Frank, M. C. (2012). Learning from others: The consequences of psychological reasoning for human learning. *Perspectives on Psychological Science*, *7*(4), 341–351.

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning* (Vol. 135). MIT Press Cambridge.

Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2017). "I won't lie, it wasn't amazing": Modeling polite indirect speech. In *Proceedings of the thirty-ninth annual conference of the Cognitive Science Society*.