Balancing informational and social goals in active learning

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

Our actions shape what we learn. Recent work suggests that people engage in efficient self-directed learning to maximize information gain. However, human learning often unfolds in social contexts where learners not only face informational goals (e.g. learn how something works) but also social goals (e.g. appear competent and impress others). How do these factors shape learners' decisions? Here, we present a computational model that integrates the value of social and information goals to predict the decisions that people will make in a simple active causal learning task. We show that an emphasis on performance or self-presentation goals leads to reduced chances of learning (E1). Next, we show that social context can push learners to pursue performance-oriented actions even when the learning goal is highlighted (E2). Our formal model of social-active learning successfully captures the empirical results. These findings are the first steps towards understanding the role of social reasoning in active learning contexts.

Keywords: active learning; social reasoning; information gain; OED; self-presentation; goal tradeoffs

Introduction

Imagine you are a novice cook and you have to decide what meal to prepare for a first date. Should you choose an easy favorite or should you attempt to make something new? While the familiar recipe can ensure a good meal, you may miss out on a new, delicious dish. The new recipe might taste even better, but it has a higher chance of failure.

We often have to choose between *exploration* and *exploitation*: that is, 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*; Sutton & Barto, 1998). This decision of whether to explore or exploit is directly related to the relative strength of the goals within a particular context. In the cooking example, I can prioritize the goal of learning by cooking the new recipe, or I can instead focus on the performance goal by preparing the tried and true meal. Here, we explore the idea of this learning-performance goal tradeoff in a simple active learning context, where social factors may shape the goals we consider.

Active learning refers to situations where people are given control over the sequence of information in a learning context (e.g. try pressing different buttons on a toy, one by one, to see if it produces an effect). 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), machine learning (Settles, 2012), and cognitive psychology (Castro et al., 2009), with the common finding that active contexts lead to faster learning than passive contexts where people don't have control over the information flow.

But real-world learning usually takes place in rich social contexts with teachers, peer learners, or other people who can directly influence our learning. Indeed, it has been suggested that children and adults modulate their inferences depending on whether they generate their own evidence, or learn from evidence generated by others (e.g. Xu & Tenenbaum, 2007). Other work suggests that children learn faster when observing intentional (more informative) actions compared to accidental (less informative) actions (Carpenter, Akhtar, & Tomasello, 1998). 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).

However, social influences are not only present when we learn from others. Even when we learn from our own actions, our social environment may affect our self-directed learning process. While previous models have captured how we optimize learning, either from our own actions or from others, they have been agnostic to other social factors that are ubiquitous in a learner's environment. People must integrate the value of social goals (e.g. looking competent or knowledgeable) and information goals when deciding what to do next. Moreover, actions that maximize learning are inherently risky in that you can potentially fail to produce an immediate outcome, and thus may be more difficult to undertake with someone else present who might judge you as incompetent.

How can active learning models accommodate this richer set of utilities? As a step towards answering this question, we model a learner who considers a mixture of learning and performance goals. A key assumption underlying inferences in recent Bayesian models of human social cognition is that people expect others to 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 a utility function that represents a weighted combination of multiple goals (Yoon, Tessler, Goodman, & Frank, 2017). Our model thus reflects a tradeoff between different goals that a learner has in a social learning context.

We instantiate our model in a simple causal learning task and examine whether people choose actions that support learning vs. social goals. We present a toy with an uncertain causal mechanism (Figure 1), for which doing only one of the two possible actions (handle pull or button press) is disambiguating but potentially risks no immediate effect (i.e. neither sound nor light turning on), while doing both actions simultaneously is immediately rewarding but is not informative

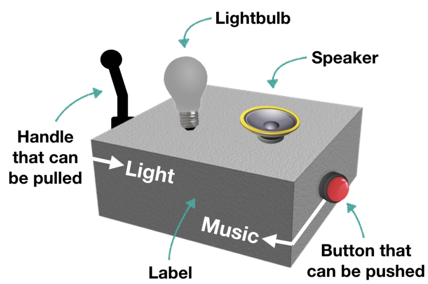

Figure 1: An example of the toy used in our paradigm.

for learning the toy's causal mechanism. Thus, the learner can choose between the two actions that will each lead to one outcome (new discovery) or the other (immediate reward). The learner's action rests on relative utilities he assigns to exploration versus exploitation, which in turn are determined in part by the presence or absence of another person he cares about (i.e. his boss).[1]

In two experiments, we show that emphasizing performance or self-presentation goals leads to actions that are not informative and thus reduce the chances of learning (E1). Next, we show that the presence of an observer (i.e., a boss) pushes learners to pursue performance/presentation actions even when the learning goal is highlighted (E2). Finally, we present a Bayesian Data Analysis showing that the empirical results are consistent with predictions of our cognitive model of social-active learning.

Computational model

We model a learner's action choice 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 2).

Learning utility The *learning utility* symbolizes the goal to learn new information, which in our paradigm 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). The learner considers the hypothesis space H, and wants to determine the correct hypothesis. Based on a set of queries, each realized through taking an action, the learner thinks about the utility of the answer to each query. The utility of answer is equal to the *information gain*, which is 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_{learn}):

$$U_{learn} = U(a) = \frac{ent(H) - ent(H|a)}{log_2 n}$$

where $ent(H)$ is the Shannon entropy of H, which provides a measure of the overall amount of uncertainty in the learner's beliefs about the candidate hypothesis(MacKay, 2003). Once the learner chooses a query Q, which yields an answer a, then he updates his beliefs about each hypothesis via standard Bayesian updating. Finally, the difference in entropy is normalized by $log_2 n$, where n is the number of possible actions, to generate a value between 0 and 1.

Performance utility The *performance utility* is the utility of successfully making the toy operate and achieving an immediate rewarding outcome. Within our paradigm, the learner gains utility by seeing an immediate effect of music or light turning on. The expected performance utility (U_{perf}) before the learner chooses an action is then the likelihood of an effect m given the learner's action a. Thus, the performance utility is maximized when the toy is expected to "go."

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

When there is no observer present ($obs = no$), 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:

$$U(\phi; obs = no) = \phi_{learn} \cdot U_{learn} + \phi_{perf} \cdot U_{perf},$$

where ϕ is a mixture parameter governing the extent to which the learner prioritizes learning over performance.

Presentation utility When there is another person present to observe the learner's action, this observer O is expected to reason about the competence c of the learner L which is equal to whether the learner was able to make the toy produce an effect. The learner thinks the observer's inferential process, and the expected *presentational* utility (U_{pres}) is based on maximizing the apparent competence inferred by the observer.

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

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

When there is an observer present ($obs = yes$), the learner considers the tradeoff between all three utilities: the learning utility, performance utility and presentational utility:

$$U(\phi; obs = yes) = \phi_{learn} \cdot U_{learn} + \phi_{perf} \cdot U_{perf} + \phi_{pres} \cdot U_{pres}$$

The learner L chooses his action a approximately optimally (as per optimality λ) based on the expected utility given his goal weights and observer presence.

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

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

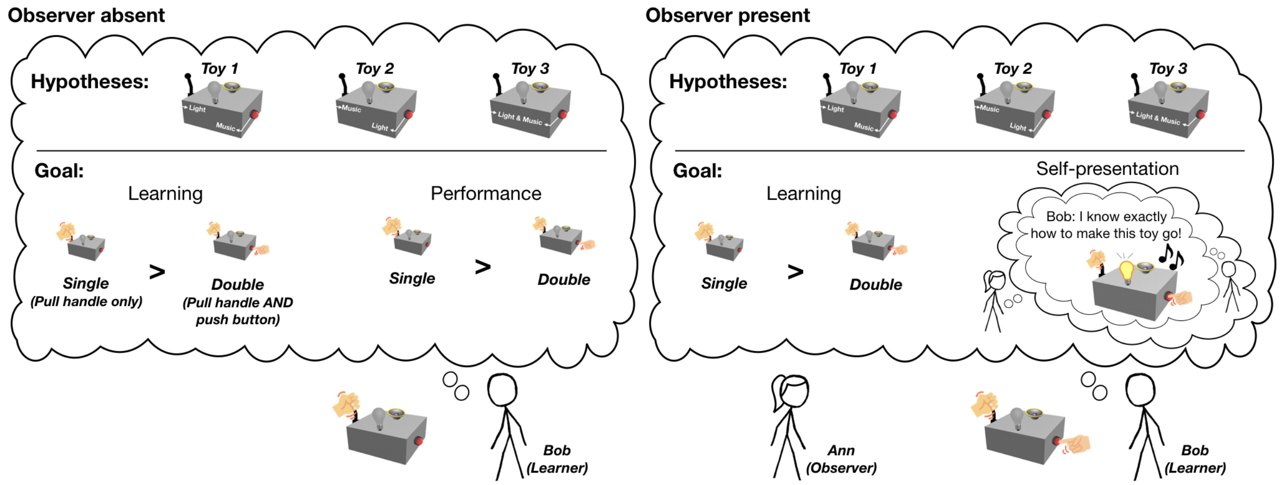

Figure 2: Diagram of the computational model. The learner considers possible hypotheses: Toy 1 (handle pull turns on the light, button press turns on music, both actions cause both effects); Toy 2 (handle pull turns on music, button press turns on the light, both actions cause both effects); and Toy 3 (both actions cause both effects, but each action on its own does not produce any effect). The learner also considers 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. The learning goal favors a single action (e.g. pull the handle only) that can fully disambiguate, whereas the performance goal favors the both action (pull the handle AND push the button) that guarantees the most salient reward. 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 or knowledge of how the toy works).

Experiment 1

In Experiment 1 (E1), we first wanted to confirm that participants would choose different actions depending on what goal was highlighted. We were also interested in how people would act when no explicit goal was specified within the task. Participants were asked to imagine that they needed to act on a toy with an uncertain causal mechanism, and were assigned to different goal conditions: (1) learning (learn how the toy works), (2) performance (make the toy play music), (3) presentation (impress their boss), and (4) no goal specified. 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.[2]

Method

Participants We recruited 196 participants (45-51 per condition) on Amazon's Mechanical Turk, with IP addresses in the 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.

[2]Our hypothesis, method, model and data analysis were preregistered prior to data collection on the Open Science Framework (https://osf.io/kcjau). All experiments, data, model scripts, and analysis codes for the statistical models can be found in the online repository for this project: https://github.com/kemacdonald/soc-info.

Stimuli and Design We presented images and instructions for three different toys that looked very similar but worked in different ways (see captions for Figure 2 2). The instructions communicated that pressing the button and pulling the handle was immediately rewarding but uninformative (fails to disambiguate the causal mechanism). In contrast, 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 the correct action(s)–outcome link.

We asked participants to act on one of these toys; importantly, the given toy was missing its label, leading to uncertainty about its causal structure. We randomly assigned participants into four goal conditions. In the *No-Goal* condition participants selected actions without any goal specified. In the *Learning*, *Performance*, and *Presentation* conditions, we asked participants to imagine that they were children's toy developers and one day their boss approached them. We then instructed participants to: figure out the correct label for the toy (*Learning*); make the toy play music (or turn the light on; *Performance*); or impress their boss and show that they are competent (*Presentation*). We asked participants to select an action out of the following set: "press the button", "pull the handle", or "press the button and pull the handle." The order of actions was randomized.

Procedure In the *exposure phase*, we showed participants an example toy and gave instructions for three toy types. We first presented the instructions for the single action toys (Toy

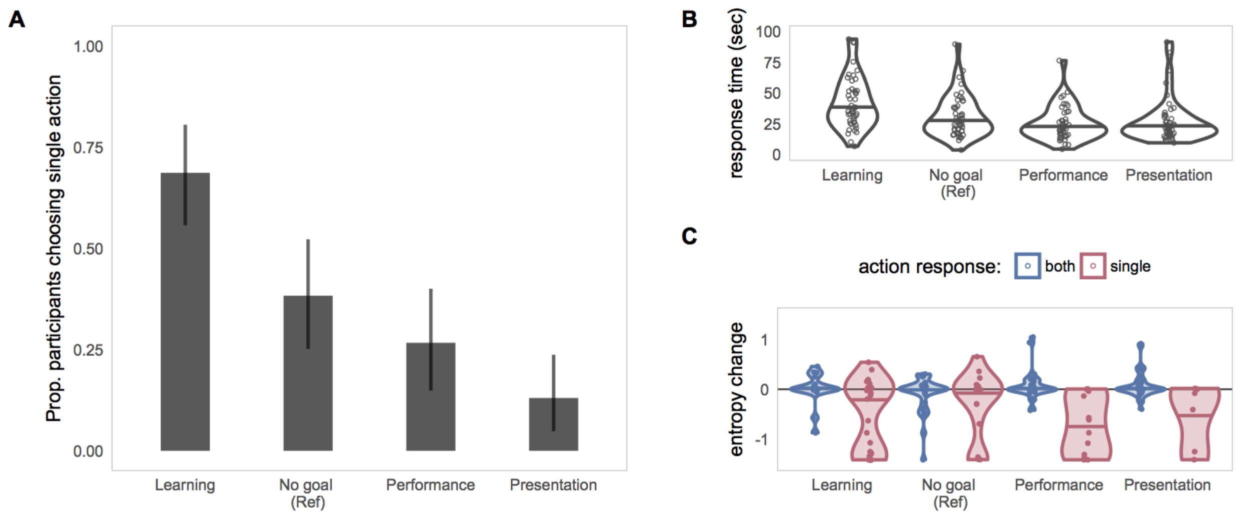

Figure 3: Behavioral results for E1. Panel A shows the proportion of action decisions for each goal condition. Error bars represent 95% binomial confidence intervals computed using a Bayesian beta-binomial model. Panel B shows violin plots of participants' response times on the action decisions. Each 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; information gain in bits) 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.

1 and Toy 2) in a randomized order, and then presented the instructions for the both action toy (Toy 3). After instructions, participants indicated what action would make each toy operate (e.g. "How would you make [this] toy play music?") to show that they understood how the different toys worked.

In the *test phase*, participants 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 ... if you [achieve the given goal]".

Both before and after the critical action decision trial, we asked participants to rate the likelihood that an unknown toy was Toy 1, 2, or 3 by adjusting a continuous slider bar which ranged on a scale from "very unlikely" to "very likely." These measurements indexed participants' prior beliefs over hypotheses about how the toys were likely to function and their *belief change* after selecting an action and observing its effect.

Results and discussion

Analysis plan We present behavioral analyses of participants' (1) action decisions, (2) action decision times, and (3) belief change (i.e., learning). 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.

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 Figure 3), with the highest proportion occuring in the Learning context, followed by the No Goal context, then Performance, and the fewest single actions in the Presentation condition.

Compared to the No-Goal condition, participants selected the single action at a greater rate in the Learning condition (β = 1.28, [0.5, 2.17]) and at lower rate in the Presentation context (β = -1.41, [-2.47, -0.4]), with the null value of zero difference condition falling well outside the 95% HDI, and at similar rate in the Performance condition (β = -0.53, [-1.43, 0.35]) with the 95% HDI including the null.

TODO_KM: add pairwise comparisons of preformance/presentation and performance/no-goal.

Action decision times: We analyzed response times in log space using the same model specification. Figure 3A shows the full RT data distribution. Compared to the No-Goal condition (M= 31 seconds), participants took on on average 12.2 (4.2, 20) seconds longer to generate a decision in the Learning condition. In contrast, participants in the Performance and Presentation conditions produced similar decision times.

Belief change: We modeled change in entropy as a function of goal condition and participants' action choices:

$entropy\_change \sim goal\_condition + action\_response$ (see Figure 3C). Across all conditions, people who selected the single action showed a greater reduction in entropy (β = -0.49, [-0.64, -0.33], i.e., learned more from their action. We did not see evidence of an interaction between goal condition and action selection. However, recall that a larger proportion of participants selected the single action in the Learning context, so the probability of learning is higher in this scenario.

Experiment 2

In E1, we confirmed that participants selected different actions depending on the type of goal emphasized by the social context. In E2, our goals were three-fold: (1) to replicate the results from E1; (2) to manipulate goals *and* the presence/absence of another person indepently, allowing us to measure the interaction between goals and social context; and (3) to compare empirical data with predictions of our computational model. Our key behavioral prediction was an interaction: that participants would be less likely to select the single (more informative) action in the Learning goal and No-Goal conditions when their boss was present. We also predicted a null result: that the presence of the boss should not affect action decisions in the Performance condition.

Method

Participants Using the same recruitment and exclusion criteria as E1, we recruited 347 participants (42-51 per condition), and 325 participants were included in our final analysis.

Stimuli and Design The stimuli and design were identical to E1, except we had seven different goal × social conditions. Goals remained identical to E1; social context varied depending on whether the boss was present in the cover story (*social*) or she was absent (*no-social*). Thus, the seven conditions were: *social-learning*, *social-performance*, *social-presentation*, *no-social-no-goal*, *no-social-learning*, *no-social-performance*, and *social-no-goal*. Note that we did not have *no-social-presentation* condition, because the presentation goal was defined by presenting oneself as competent to another person.
Procedure
The procedure was identical to E1.

Results and discussion

Action decisions: We modeled action decisions using a logistic regression specified as $action \sim goal\_condition * social\_context$ with the no-goal-no-social condition as the reference category. We replicated the key finding from E1: participants selected a "single" action more often when they were in a context that emphasized a learning goal, followed by the no-goal, performance, and presentation conditions (see panel A Fig 4). There was a main effect of social context, with participants being less likely to select the single action when their boss was present (β = -0.521, [-1.005, -0.053]). Finally, there was evidence for a reliable interaction between goal condition and social context such that the effect of social context was present in the Learning and No-Goal conditions, but not in the Performance condition (β_{int} = 1.163, [0.01, 2.312]).

Action decision times: We replicated the key decision time finding from E1, with participants making slower decisions in the Learning context as compared to Performance/Presentation. On average, participants took seconds to generate a response in the No-goal condidition and seconds in the Learning condition. In contrast, decisions were faster in the Performance (β = -7.78 sec, [-14.01, -1.52]) and Presentation (-10.77 seconds, [-18.67, -2.73]) conditions, which were similar to one another (see Panel B of Fig 4). There was no evidence of a main effect of social context or an interaction between goal condition and social context. Note that here we did not see a difference in decision times between the Learning and No-Goal conditions, which is different from the pattern in E1.

Belief change: Across all conditions, participants who selected the single action showed a greater reduction in entropy (β = -0.35, [-0.45, -0.24]. There was some (weaker) evidence of greater reduction in entropy in the Learning goal condition (β = -0.12, [-0.25, 0.01). There was no evidence of a main effect of social context and no two- or three-way interactions between social context, goal condition, and type of action choice.

BDA model-data fit: In our paradigm, participants were instructed to choose an action[3] based on a certain goal. We assumed that the goal descriptions (e.g. "figure out the correct label for the toy") conveyed to the participants a particular set of goal weights $\{\phi_{learn}, \phi_{perf}, \phi_{pres}\}$ that they used to generate their action choices. We put uninformative priors on these weights ($\phi \sim Uniform(0,1)$) and inferred their credible values separately for each pair of different goal condition and social context, using Bayesian data analytic techniques (Lee & Wagenmakers, 2014).

The inferred goal weights were consistent with what we predicted (see Figure 4, panel D). ϕ_{learn} was at its highest for no-social learning condition, in which the goal to learn was highlighted, and there was minimum social pressure. On the other hand, the ϕ_{perf} and ϕ_{pres} together make up the highest portion in the presentation condition, with high social pressure to present competence, compared to other conditions.

We also inferred another parameter of the cognitive model, the optimality parameter λ. We put uninformative prior on the parameter ($\lambda \sim Uniform(0,10)$ and inferred its posterior credible value from the data. We ran 4 MCMC chains for 100,000 iterations, discarding the first 50,000 for burnin. The Maximum A- Posteriori (MAP) estimate and 95% Highest

[3]For action priors, we used a separate prior elicitation task, in which people indicated the likelihood for selecting an action without any background information about possible hypotheses or goals. The results suggested that none of the action choice priors statistically differed from chance. We used mean likelihood for each action choice as baseline priors in our model.

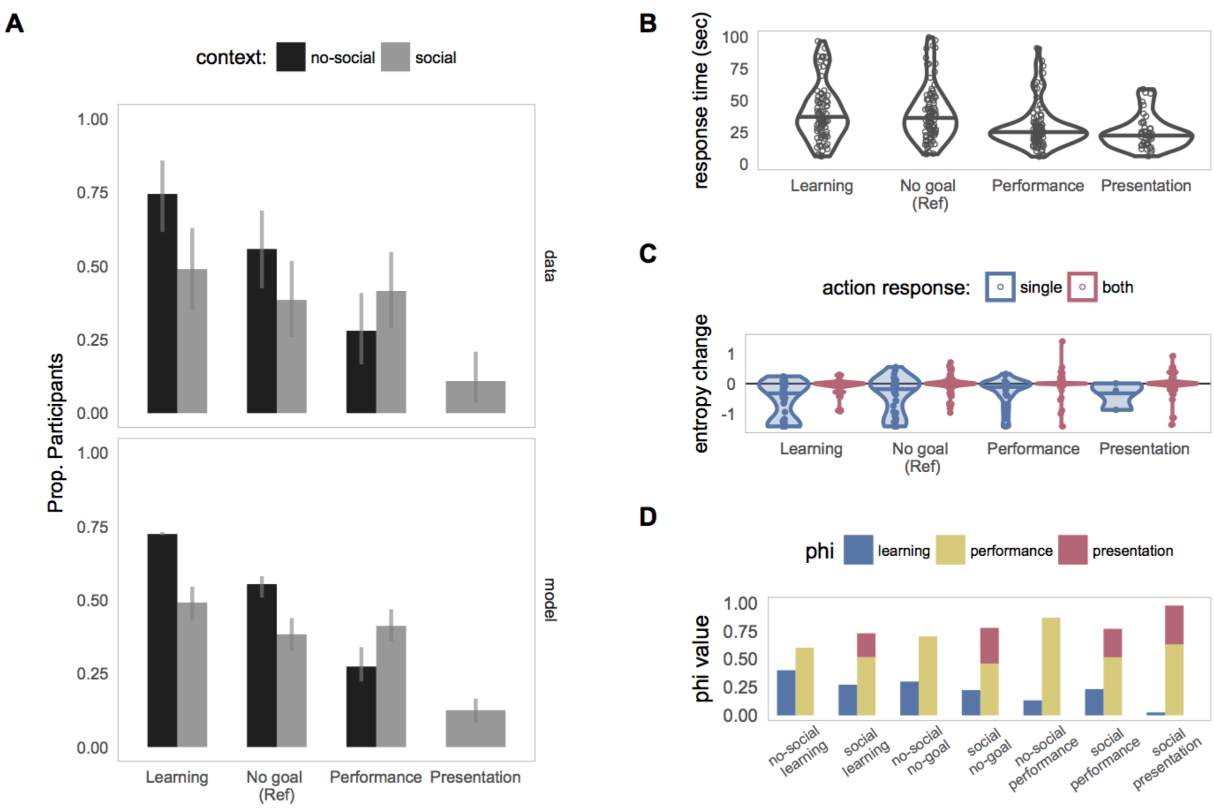

Figure 4: Behavioral and model fitting results for E2. Panel A shows actions decisions with color representing social context, from human data (top) and fitted model predictions (bottom). Panel B shows decision times. Panel C shows belief change. Panel D shows inferred phi values for each goal-context condition. All other plotting conventions are the same as Figure 3.

Probability Density Interval (HDI) for λ was 4.79 [3.96, 6.2].

The predictions of the action choices according to the fitted learner model are shown in Figure 4, panel A (bottom). The model's expected posteriors over action choices capture key differences between conditions: the single action was more likely for no-social than social conditions overall, but not when the performance goal was highlighted. The model was able to predict the distribution of action responses with high accuracy $r^2(21) = 0.9$.

General Discussion

How do social contexts shape active learning? Here, we proposed that people integrate learning-, performance-, and presentation-oriented goals when deciding what to do next. In two experiments, people's we showed that people chose more informative actions when learning goals were highlighted and in the absence of a relevant social context (i.e., no boss present), while they chose more immediately rewarding actions when performance or presentational goals were highlighted, especially when a boss was present. When no explicit goal was specified, people showed behavior that seemed to reflect a mixture of goals. Our model of social-active learning successfully captured key patterns in the behavioral data.

This work represents a way to bring active learning accounts into contact with social learning theories. We used ideas from Optimal Experiment Design, which models active learning as a process of rational choice that maximizes utility withrespect to gaining information, and Rational Speech Act theory, which formalizes a process of recursive social reasoning in language use. This step allowed us to include social information within a formal utility-theoretic framework, building a richer utility function that represented a weighted combination of multiple goals – both informational and social.

There are important limitations to this work that present opportunities for future work. First, we did not differentiate between performance and presentation goals in the current model/pardigm. That is, the choice of doing both actions satisfies both performance and presentational gaols. Our future work is aimed at enriching the space of actions that people could take, which should allow us to tease apart actions driven self-presentation. Second, we used a very particular social context – the presence of a boss – to influence people's action choices. Thus, it remains an open question as to how these results would generalize to other kinds of observers that hold different goals. One particularly compelling contrast would be to a teacher who wants the learner to select actions that help her learn. Third, we limited people to a single action choice. While this allowed us to get a clean measurement of our goal and social context manipulations, real-world learning often involves a process of sequential decision-making that could cause learners to priortize different goals depending on their own prior actions and/or the proability of interacting with an observer in the future.

Another interesting open question is how our model/results could be used to understand active learning over development. Our framework would in principle allow us to measure changes in children's goal preferences as they develop more sophisticated social reasoning and better meta-cognitive abilities. One prediction is that young children focus on learning goals earlier in development when they are surrounded by familiar caregivers who scaffold learning-relevant actions. But as their social reasoning abilities mature and their social environments become more complex, children may start to emphasize performance or presentation goals.

Overall, this work represents a first step to answering these rich questions that ultimately seek to unify theories on active learning and social reasoning.

Acknowledgements

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

References

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, *21*(2), 315–330.

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).

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

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.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.

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).

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.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as bayesian inference. *Psychological Review*, *114*(2), 245.

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*.