# Exploring Teaching with Evaluative Feedback

**Arunima Sarin (asarin@g.harvard.edu)**
Department of Psychology, Harvard University
William James Hall, 33 Kirkland Street, Cambridge, MA 02138 USA

**Fiery Cushman (cushman@fas.harvard.edu)**
Department of Psychology, Harvard University
William James Hall, 33 Kirkland Street, Cambridge, MA 02138 USA

**Abstract**

In this experiment, we explore how teachers use evaluative feedback—such as praise and criticism, or reward and punishment—to guide learners' behavior. Although common in daily life, there has been limited research in this area. Our study combines insights from Bayesian models of pedagogy and prior experimental research on evaluative feedback to address this gap. We defined an objective within a complex conceptual space and observe how teachers use only evaluative feedback to guide naive learners' choice. Our findings indicate that teachers tend to structure their feedback communicatively, in a way that minimizes uncertainty and prioritizes establishing common ground. Our results offer preliminary but exciting insights into how humans teach with evaluative feedback, providing a more comprehensive understanding of the ease and agility with which we engage in intuitive teaching.

**Keywords:** intuitive pedagogy; evaluative feedback; teaching; Bayesian inference

## Introduction

Humans are uniquely adept at pedagogy, the capacity for teaching and learning from others (Csibra, 2007; Csibra & Gergely, 2006; Tomasello, 2009). Pedagogy involves the intentional transfer of information and skills from a knowledgeable individual, the teacher, to an uninformed individual, the learner. Humans engage in pedagogy naturally and intuitively, imparting knowledge acquired through personal experience without requiring formal education (Ashley & Tomasello, 1998; Knudsen & Liszkowski, 2013).

Current research suggests that much of human pedagogy is consistent with principles of Bayesian inference, a framework that formalizes rational belief updating based on observed data. Specifically, models of recursive Bayesian inference reveal how pedagogy can be structured as a collaborative process between a teacher and a learner, where both parties are actively engaged in representing each other's mental states (Bonawitz et al., 2011; Gweon, 2021). The framework has been used with success in understanding and predicting teaching behaviors across various goals and strategies. However, most prior work focuses situations where teaching is initiated by the teacher, for instance by indicating relevant examples of a concept, demonstrating the use of a tool (Ho et al., 2016; Shafto et al., 2014). Less attention has been paid to pedagogical interactions where the learning is self-directed, such as those facilitated solely through evaluative feedback.

In pedagogical interactions of this type a learner acts, and the teacher responds with actual, verbal, or symbolic reward or punishment. Although it has not been studied much, this kind of pedagogy is common: we often use evaluative feedback to teach, and the simplicity of the feedback makes it an effective tool to teach adults, children, pets, and even robots (Fehr & Gächter, 2002; Isbell & Shelton, 2001; Owen et al., 2012). For instance, the simple phrase "good job!" can serve as an effective motivator for an adult who has completed their first coding project, a child who has finished their meal, and a pet that has successfully followed a command. The basic principle of this pedagogical interaction is also evident in simple games such as "hot-or-cold," where a knowledgeable player, who knows the location of the target object, guides a naive player's search by providing only "hot" or "cold" feedback.

But how does the knowledgeable player decide when to say 'hot' versus 'cold'? Our study aims to understand conceptually how human teachers provide evaluative feedback. To do so, we draw upon key insights from two distinct literatures: one that demonstrates the usefulness of Bayesian inference in understanding human pedagogy, and another on the use of evaluative feedback by humans. We briefly cover these key insights in the following sections.

## Pedagogy & Bayesian Inference

Bayesian inference is a framework for probabilistic reasoning under uncertainty. It comprises of two key ideas. First, it posits that beliefs are probabilistic representations rather than point estimates—for instance, we represent the likelihood of rain as a distribution rather than a single setting of a binary variable ("will" or "won't"). Secondly, it provides a formal model for updating beliefs in response to new information through the application of Bayes' theorem. The framework has been applied successfully to model various facets of human cognition, including teacher-driven pedagogy (Ho et al., 2018; Oaksford & Chater, 2009).

One such application examined how teachers imparted a rule-based concept, specifically the boundary of a rectangle,



In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.* 

by providing examples of points within on a blank screen to a learner (Shafto et al., 2014). Rather than selecting examples in a random or arbitrary manner (selecting any point on the screen or any point within the rectangle), teachers consistently chose examples that were maximally informative, such as two positive examples at opposite ends of the rectangle, or one positive one negative example side-by-side. This pattern of example selection was in line with the predictions of model of pedagogy that implements recursive Bayesian inference. According to the model, teachers represent what a learner is likely to infer given various examples, and they therefore select examples that maximize a learner's belief in the correct concept. The learners' inferences are dependent on their prior beliefs and the degree to which the examples are likely to be chosen by a helpful teacher. This mutual dependence between teacher and learner is a key conceptual insight into the way in which a knowledgeable teacher designs instruction. Many instances of teacher-led pedagogy operate in this manner (Ho et al., 2021; Popp & Gureckis, 2020). However, it remains to be determined whether these conceptual insights also apply to learner-led pedagogy, such as in cases where teachers are limited to providing evaluative feedback.

## Teaching with Evaluative Feedback

A simple way for us to teach others is by providing them with positive feedback when they do something good and negative feedback when they do something bad. One well-studied example is teaching with rewards and punishments. In principle, teaching with and learning from rewards and punishments could unfold in two ways: teachers and learners could use the feedback as reinforcements (meant to reinforce local action-outcome pairs), or they could engage in theory of mind reasoning and use the feedback communicatively (to communicate about the general appropriateness of an action in achieving the desired goal). Current research suggests that teachers and learners often rely on the latter strategy, provisioning evaluative feedback communicatively and interpreting it as such. For instance, when provided with feedback in the form of praise or critique, learners often tend to draw sharply different inferences about their competence based on whether the feedback came from a teacher who had prior knowledge of their ability or one who did not (Barker & Graham, 1987; Meyer, 1982, 1992; Miller & Hom, 1997). On the other hand, when faced with deciding how to punish a transgressor, teachers often select punishments that prioritize the message they want to communicate through their punishment, even if it means imposing less harsh penalties (Molnar et al., 2022; Sarin et al., 2021), including foregoing the punishment altogether if the message cannot be interpreted accurately due to ambiguity in the circumstances (Rai, 2022). Teachers/punishers also report being satisfied with punishment only if the learner/transgressor signals that they have understood it "as a message" and not simply experienced it as a negative incentive (Funk et al., 2014; Gollwitzer & Denzler, 2009). These findings seem consistent with the idea that when people are deciding what feedback to give, or what inferences to draw from the feedback they have received, they draw upon an explicit model of communicative intent.

Ho et al., (2018) tested this idea directly. Participants in their studies were tasked with teaching a virtual dog to walk along a path using only evaluative feedback. Across a series of experiments, dogs programmed to treat human feedback as communication learnt the target path faster and more efficiently than dogs programmed to maximize human rewards, suggesting that people tend to use rewards and punishments as a form of communication (relying on our capacity for theory of mind) rather than as reinforcement (relying on our capacity for simple reward learning).

Their findings offer valuable insight into how people approach teaching with evaluative feedback. However, they do so by characterizing and comparing two classes of *learner* models: reward-maximizing and communicative. As such, it leaves open the question of how *teachers* solve the challenge of structuring their feedback. This is the focus of our research.

## The Present Study

We take as our starting point the finding that teachers use evaluative feedback, such as rewards and punishments, communicatively rather than just as reinforcements. The question that arises, then, is what kind of cognitive strategy they use to accomplish this.

We designed an experiment aimed at understanding how teachers effectively guide novice learners in identifying a complex concept. Participants, who played the role of teachers, were shown a set of objects, each made up of six features (shape, color, pattern, center, tone, and boundary) that could take on one of two possible values, yielding 64 unique configurations. Each object was tied to a specific reward, with one object possessing the highest reward, and serving as the target concept. While the teachers knew the identity of the target concept, the learners did not. The teachers' task was to assist the learners in identifying the target object solely through evaluative feedback, which was given on a scale of very bad, bad, neutral, good, and very good, as learners made a series of choices.

## Experiment

Our experimental setup is designed to explore how teachers structure evaluative feedback from trial trial-to-trial. We consider three potential strategies. These strategies are neither mutually exclusive, as teachers may choose to utilize them in conjunction, nor are they exhaustive, as there may exist additional ways to solve this task. However, they cover a rich conceptual space, and serve as a valuable starting point for an initial investigation into this topic.

The simplest approach for teachers is to *mirror* the reward function of the environment (which, again, is privately known to them but not the learner). This is a cognitively simple strategy, and, in the long run, it should impart an accurate understanding of the task to the learner.

Next, we considered a *heuristic* strategy in which the teacher constructs a monotonic gradient of reward defined by proximity, in feature space, to the target concept. In other words, the closer the learner is to selecting the target concept, the more reward the teacher provides.

Finally, we considered the possibility that teachers use a mental model of the learner to select the specific feedback that maximizes the learner's likelihood of subsequently choosing the target concept. This *belief-directed* strategy requires the teacher to consider the learner's mental state and the inferences they are likely to make based on the feedback received. The teacher must then select feedback that adjusts the learner's (probabilistically represented) belief towards the target concept. Giving feedback in this way would require the teacher to rely on their capacity for theory of mind and would be conceptually similar to models of human pedagogy as a form of recursive Bayesian inference.

## Methods

**Participants** 245 adult residents of the U.S. were recruited from Amazon's Mechanical Turk to take part in the experiment. After removing those who left the study incomplete, the final sample was made up of 243 participants ($M_{\text{agegroup}}$ = 25-44, 44% female identifying).

**Materials and Design** To understand how teachers structure their rewards and punishments to help a naïve learner learn a target concept, we created an experiment involving 64 complex objects. Each object was made up of 6 features, each of which could take on one of two levels - color (red or purple), shape (square or squiggle), boundary (gray or plain), pattern (solid or stripes), tone (one-tone or two-tone), and center (empty or filled). This gave us a total of 64 unique objects (see Figure 1).

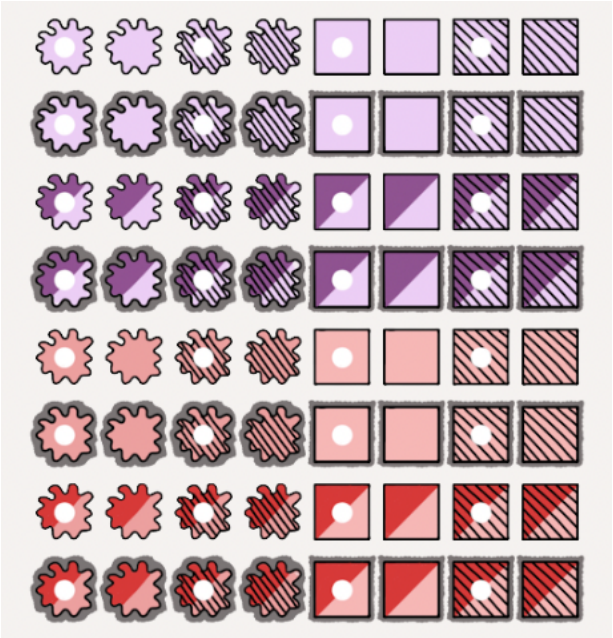

Figure 1: Complex object stimuli used in Experiment 1.

One of these objects was randomly selected to be the target object and assigned a value of 100 points. All other objects were assigned a value of 1 point each. Participants, who assumed the role of teachers, were informed that their task was to maximize the number of points earned. The most efficient means of achieving this was to select the target stimulus. However, teachers were not allowed to make selections directly; instead, they were informed that they would be paired with a "partner" (referred to henceforth as the "learner"), who could only see the set of stimuli on their screen and not the associated rewards. The task of the teacher, therefore, was to assist the learner in earning the highest possible number of points, presumably by identifying the target concept. On each trial, teachers were presented with the stimulus chosen by the learner and the target stimulus and were required to provide evaluative feedback in the form of very bad, bad, neutral, good, or very good using the scale shown in Figure 2, for a total of 14 trials.

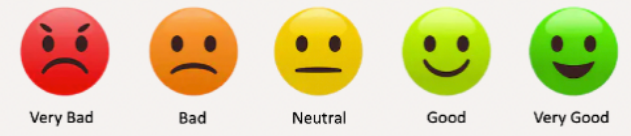


Figure 2: Evaluative feedback scale given to participants.

To create conditions that would allow for comparisons of the different teaching strategies, we programmed three distinct learner behaviors. All learners behaved the same in the first three trials of the experiment. Specifically, on trial 1, learners chose an object that had nothing in common with the target. On trial 2, they chose an object with exactly one feature in common with the target (always shape) and on trial 3, they chose an object with two features in common with the target (always shape & color). Then, on trial 4, all three learners made different selections. The *novel learner* chose an object that shared two new, yet untested feature-settings with the target (e.g., pattern and tone). The *incidental learner* selected an object that shared one new feature with the target (e.g., pattern) and a feature they happened to get right on just the previous trial (i.e., color). Finally, the *base learner* selected an object with one new feature (e.g., pattern) and a feature they had systematically gotten correct in the last two trials (i.e., shape) (see Figure 3). It is worth noting that all the three learners always selected objects that shared two features with the target. The distinction in their selections lay only in the specific features they identified correctly, with some selections consisting of previously tested and acknowledged features, while others comprising of new, untested features.

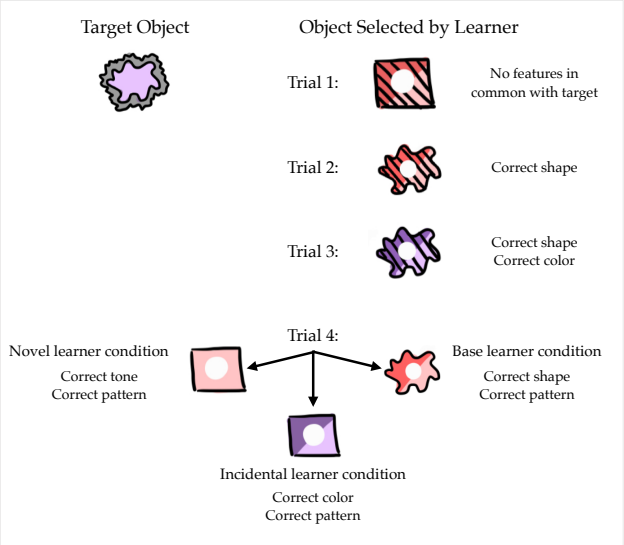


Figure 3: Visual representation of objects selected by the three types of learners.

Given the experimental design, we can draw out qualitative predictions for the three strategies a teacher may employ to structure their feedback on trial 4. By comparing the actual behavior of the teacher in the task with these predictions, we can gain a preliminary understanding of how teachers might approach tasks that are exclusively dependent on evaluative feedback.

A teacher using the *mirroring* strategy should provide the same feedback to all three learners. This is because the object selected by each learner falls outside the target object by the same number of features (i.e., 4) and is worth the same number of points (i.e., 1). So, regardless of whether the teacher is mirroring their feedback with the reward structure of the environment or the conceptual boundary between the target and non-target objects, their feedback should remain unchanged across different learners. A teacher employing a *heuristic* to structure their feedback should also treat the three learners similarly. For a teacher of this type what determines the feedback is the chosen object's distance from the target in feature space, irrespective of the type of features selected. Since all three learners pick objects that share exactly 2 features with the target, they should receive the same feedback from this teacher. Finally, a teacher relying on a *belief-directed* strategy should treat the three learners differently. This is because their feedback depends on the inference they make about the hypothesis the learner is entertaining from the learners' selections. As each of the three learners exhibits different patterns of selections, a belief-directed teacher would draw different inferences for each learner, thereby resulting in different feedback aimed at altering their beliefs towards the target concept.

A *belief-directed* teacher could attain this objective in two potential ways. One approach would be to *maximize informational gain* for the learner on every trial. This would entail rewarding the learner whenever they accurately associate a feature-setting with the target, and more so when they correctly identify new features. A teacher following the informational gain strategy would thus offer the most positive feedback to the novel learner due to their correct identification of two new features, followed by the incidental and base learners. This strategy is the most efficient in principle, however recent research indicates that teachers may adopt a different approach in practice (Popp & Gureckis, 2020). A challenge with this approach is the teacher's lack of certainty regarding the learner's inference and cognitive ability, making it difficult for them to effectively shape the learner's beliefs without a clear understanding of the learner's knowledge and thought process. The teacher thus has to rely on limited evaluative feedback to influence the learner's belief under a significant degree of uncertainty. So, a different approach teachers could take would be to use their feedback to first *establish a common ground* between themselves and the learner and then use it as a scaffold to impact information successively. Although suboptimal compared to the informational gain approach, this approach would allow a teacher to reduce their uncertainty about the learner's inferences. Recent studies indicate that this is indeed something people tend to do, as when teachers are faced with the task of balancing asking questions and providing instructions, they tend to ask more questions than what would be predicted by an optimal model (Popp & Gureckis, 2020), implying a tendency to engage in suboptimal behavior to reduce their uncertainty before imparting information (Bradac, 2001; Epstein, 1999). Thus, a teacher following this approach in our task would provide the most positive feedback to the base learner and the most negative feedback to the novel learner, as the latter's correct identification of two new features comes at the expense of gains established through collaboration with the teacher.

## Results

To examine the relationship between participant feedback and different types of learners, we transformed the evaluative feedback scale into an ordinal scale, with a range of -2 to 2. We then regressed participants feedback on trial 4 (the critical trial) on the different learner types using a linear model. To account for the various possible hypotheses, we created two dummy codes, one for the base learner and another for the novel learner, making incidental learner our reference category. Results reveal that the model was a significant fit ($F(2, 240) = 24.91$, $p < .001$). Overall, all three learners received negative feedback (see Figure 4)[1]. Compared to our reference category, the incidental learners ($M = -0.93$, $b = -0.93$, $p < .001$), base leaners received more positive feedback ($M = -0.27$, $b = 0.65$, $p < .001$) while novel learners received more negative feedback ($M = -1.32$, $b = -0.4$, $p = .009$). These results fall in line with the qualitative predictions of a teacher

[1] To increase precision and check for robustness, we ran a second model adding participants own feedback on the trial 3 as a predictor along with the learner types. The direction and significance of the estimates) remain qualitatively similar.

structing their feedback to shape the learner's belief using a strategy to minimize uncertainty by establishing common ground.

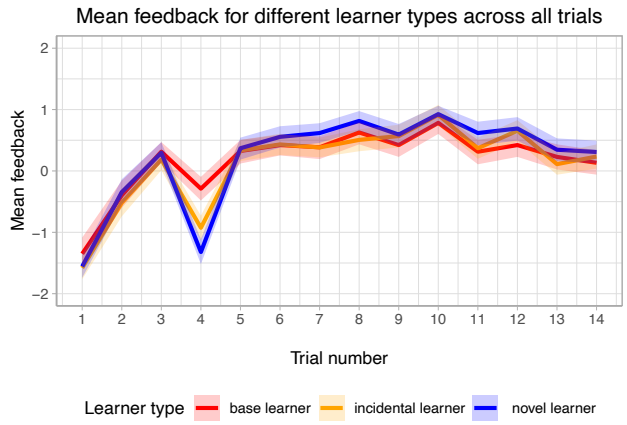


Figure 4: Mean feedback for each learner type.

## Discussion

We explore how people teach intuitively using rewards and punishments. Teaching in this way is prevalent, yet little research exists on this subject. In our experiment, participants, playing as teachers, were shown a set of complex objects. Each object was made up of one of two settings of 6 features and one of the one object was more valuable than others. Teachers were paired with learners who they were told could see the objects but not their reward and tasked with helping the learner identify the target object using only evaluative feedback.

To understand how teachers structure their feedback, we examined three potential strategies: mirroring the environment, utilizing a heuristic approach, or engaging in a form belief-directed pedagogy, using an approach to either maximize informational gain through feedback or establish common ground through feedback. The results of our experiments demonstrate that neither mirroring nor heuristic strategies are employed alone. For example, as the learner progresses from correctly identifying no features to correctly identifying one feature, the average feedback improves, indicating that mirroring alone is not being utilized, as the underlying reward function is unchanged in this region. Furthermore, the results from the critical trial indicate that a simple heuristic strategy, which focuses on minimizing the distance between the current selection and the target in feature space, is not being utilized as all three types of learners—the base learner, incidental learner, and novel learner—should have received the same feedback. After all, each of these learners selects an object that shares exactly two correct features with the target. However, these three learners were treated differently by participants in our study.

Our findings suggest that teachers tend to employ principles of belief-directed reasoning, utilizing feedback strategically to influence the learner's understanding given a model of the learning process. Additionally, the pattern of results also provides insight into the nature and structure of this pedagogy. A teacher could aim to shape a learner's belief in at least two ways: by maximizing informational gain across all features simultaneously for each trial, or by reducing their own uncertainty about the learner's inferences by establishing common ground and then imparting information about features successively. Our results provide evidence for the latter approach. We find that a novel learner receives the most negative feedback as they give up the two features they had previously gained with the teacher.

Our findings may seem surprising at first—after all, teacher's strategy of minimizing their own uncertainty is suboptimal compared to the strategy that prioritizes informational gain for the learner. Nevertheless, recent studies suggest that human reasoning may, in fact, exhibit suboptimal behavior along similar lines (Popp & Gureckis, 2020). It is also worth noting that the complexity of the stimuli we used in this experiment, which consisted of six different features, may have contributed to the adoption of the cognitively simpler strategy of establishing common ground. It's therefore an open question whether people's adoption of this strategy is specific to the current experimental design or if it reflects a more general tendency of how people teach with evaluative feedback. To that end, a fruitful next step would be to simplify the feature space and see if there is a shift in teacher strategy.

Bearing this caveat in mind, our results suggest that people may structure their evaluative feedback communicatively, using their capacity for recursive theory of mind. On this view, teachers have a mental representation of the learning process, and they use this to identify the specific evaluative feedback that will maximize the probability of the learner choosing the target concept on the next round. Meanwhile, we propose that learners likely recognize the teacher's pedagogical communicative intent and learn from the feedback accordingly (Although learner behavior was not directly explored in our study). Mutual inference of this kind can be best modeled as recursive Bayesian inference. Building a formal computational model remains an important area for future work.

Our study provides a valuable initial exploration into an under-researched area. However, there are some caveats. In this iteration of the experiment, we manipulated the behavior of the learners so that they selected the same correct features on trials 1-3. Specifically, all learners selected shape as the base feature and color as the incidental feature. This enabled us to tightly control the behavior of the learners and measure differences solely in teacher response. However, features such as color and shape may be more salient and easier to communicate about than features such as pattern or tone. Therefore, future work must expand upon these findings by utilizing different features as the base and incidental features. Additionally, in the experiment, 63 of the shapes had 1 point, while the target had 100 points, mimicking a needle-in-a-haystack scenario. While this approach simplified the experimental setup, it would be beneficial to introduce a varied reward distribution in future studies. This would allow

us to observe how individuals may trade-off information, such as closeness in feature space with low or negative reward.

## Acknowledgments

We thank Carter Allen for their help with data collection, Adam Bear for their helpful discussion of the concept and strategies, and the reviewers for their feedback. This research was funded by Award N00014-22-1-2205 from the Office of Naval Research to FC.

## References

Ashley, J., & Tomasello, M. (1998). Cooperative Problem-Solving and Teaching in Preschoolers. *Social Development*, *7*(2), 143–163.

Barker, G. P., & Graham, S. (1987). Developmental study of praise and blame as attributional cues. *Journal of Educational Psychology*, *79*, 62–66.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322–330.

Bradac, J. J. (2001). Theory Comparison: Uncertainty Reduction, Problematic Integration, Uncertainty Management, and Other Curious Constructs. *Journal of Communication*, *51*(3), 456–476.

Csibra, G. (2007). Teachers in the wild. *Trends in Cognitive Sciences*, *11*(3), 95–96.

Csibra, G., & Gergely, G. (2006). *Social learning and social cognition: The case for pedagogy*.

Epstein, L. G. (1999). A Definition of Uncertainty aversion. *The Review of Economic Studies*, *66*(3), 579–608.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 4.

Funk, F., McGeer, V., & Gollwitzer, M. (2014). Get the Message: Punishment Is Satisfying If the Transgressor Responds to Its Communicative Intent. *Personality and Social Psychology Bulletin*, *40*(8), 986–997.

Gollwitzer, M., & Denzler, M. (2009). What makes revenge sweet: Seeing the offender suffer or delivering a message? *Journal of Experimental Social Psychology*, *45*(4), 840–844.

Gweon, H. (2021). Inferential social learning: Cognitive foundations of human social learning and teaching. *Trends in Cognitive Sciences*, *25*(10), 896–910.

Ho, M. K., Cushman, F., Littman, M., & Austerweil, J. L. (2018). *People Teach with Rewards and Punishments as Communication not Reinforcements*. PsyArXiv.

Ho, M. K., Cushman, F., Littman, M. L., & Austerweil, J. L. (2021). Communication in action: Planning and interpreting communicative demonstrations. *Journal of Experimental Psychology. General*, *150*(11), 2246–2272.

Ho, M. K., Littman, M., MacGlashan, J., Cushman, F., & Austerweil, J. L. (2016). Showing versus doing: Teaching by demonstration. *Advances in Neural Information Processing Systems*, *29*.

Isbell, C., & Shelton, C. (2001). Cobot: A Social Reinforcement Learning Agent. *Advances in Neural Information Processing Systems*, *14*.

Knudsen, B., & Liszkowski, U. (2013). One-Year-Olds Warn Others About Negative Action Outcomes. *Journal of Cognition and Development*, *14*(3), 424–436.

Meyer, W.-U. (1982). Indirect communications about perceived ability estimates. *Journal of Educational Psychology*, *74*, 888–897.

Meyer, W.-U. (1992). Paradoxical Effects of Praise and Criticism on Perceived Ability. *European Review of Social Psychology*, *3*(1), 259–283.

Miller, A. T., & Hom, H. L. Jr. (1997). Conceptions of ability and the interpretation of praise, blame and material rewards. *Journal of Experimental Education*, *65*, 163–177.

Molnar, A., Chaudhry, S., & Loewenstein, G. (2022). "It's Not About the Money. It's About Sending a Message!": Avengers Want Offenders to Understand the Reason for Revenge. In *SSRN Electronic Journal*.

Oaksford, M., & Chater, N. (2009). Précis of *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. *Behavioral and Brain Sciences*, *32*(1), 69–84.

Owen, D. J., Slep, A. M. S., & Heyman, R. E. (2012). The Effect of Praise, Positive Nonverbal Response, Reprimand, and Negative Nonverbal Response on Child Compliance: A Systematic Review. *Clinical Child and Family Psychology Review*, *15*(4), 364–385.

Popp, P., & Gureckis, T. (2020). Ask or tell: Balancing questions and instructions in intuitive teaching. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*, 1229–1235.

Rai, T. S. (2022). Material Benefits Crowd Out Moralistic Punishment. *Psychological Science*, *33*(5), 789–797.

Sarin, A., Ho, M. K., Martin, J. W., & Cushman, F. A. (2021). Punishment is Organized around Principles of Communicative Inference. *Cognition*, *208*, 104544.

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, *71*, 55–89.

Tomasello, M. (2009). *The Cultural Origins of Human Cognition*. Harvard University Press.