Metacognitive Bandits: When Do Humans Seek AI Assistance?

Aakriti Kumar[1], Trisha Patel[2], Aaron Benjamin[2], Mark Steyvers[1]

Abstract— **Humans increasingly collaborate with AI systems to make complex decisions in the real world. While a lot of work is being done to make more accurate and interpretable AI, little is known about when and how humans decide to look towards AI assistants for help. To address this gap, we develop metacognitive bandits: a computational model of a human's advice-seeking behavior when working with an AI. The model describes a human's metacognitive process of deciding when to rely on their own judgment and when to solicit the advice of the AI based on their assessment of the utility of the AI's advice. It also accounts for the difficulty of each trial in making the decision to solicit advice. We illustrate that the metacognitive bandit makes decisions that are qualitatively similar to humans in a behavioral experiment.**

INTRODUCTION

Human decision-makers are increasingly reliant on AI assistance in domains that were previously thought to be exclusively dependent on human subjectivity and expertise [1], [2]. A common pitfall of such hybrid human-AI decision making is the ineffective treatment of advice from an AI agent by the human. To correctly assess and use an AI agent's advice, the human must infer the agent's expertise and knowledge about the task at hand, i.e, the human must employ machine theory of mind to build a mental model of the AI's ability. In this paper, we present a cognitive science perspective on how humans infer an AI agent's ability and use this inference to guide their decision to solicit the AI's advice as opposed to relying on their own judgement.

Resistance to outside advice is not unique to human-AI teams: humans discount advice from peers and tend to rely on their own judgment, even when that judgment is from an expert [3]. Humans also exhibit excessive and unwarranted confidence in their own judgments relative to those of their peers [4]. Recent work suggests that a number of similar behaviors might be at play when humans collaborate with AI which can lead to sub-optimal outcomes [5]–[7].

The human-machine interaction literature reports two contrasting biases that humans are susceptible to when working with AI: *algorithm appreciation* and *algorithm aversion*. Algorithm appreciation is the tendency of a human to prefer algorithmic help over another human's help [6]. In contrast, algorithm aversion has been described as the tendency of a human to disregard the recommendations of a machine after observing that it made a mistake. This can occur even when the algorithm can be beneficial to the human decision maker on average [7]. Human behavior consistent with these biases is often reported as inappropriate reliance by the human on the AI.

In this paper, we present a cognitive model for human-AI collaboration: we argue that varying degree of reliance on AI is a consequence of quasi-optimal decision-making on the part of the human. The human's decision to ask for help can be thought of as a metacognitive exercise - the human reflects on their own knowledge relative to the AI to make a decision on whether to seek help or not. This decision to ask for help can be formulated as a combination of two cognitive processes: *explore/exploit sequential decision-making* and *metacognition*. We start with an assumption that humans behave like quasi-ideal observers, performing Bayesian inference to decide when to ask for AI assistance. We posit that humans engage metacognition to infer and compare the utility of making one's own decision with the utility of seeking the advice of an AI. This relative assessment guides the decision to seek advice of the AI or rely on their own judgement. We model the sequential decision-making problem of soliciting advice on each trial as an explore/exploit problem. The human can explore by choosing to solicit the advice of the AI. This action is risky, since the AI has an unknown capability and the action to solicit advice is associated with time costs associated with soliciting, processing, and integrating the advice with one's own judgment. The decision to seek advice pays off if the utility of AI advice exceeds the utility of making an independent decision. The AI's expertise can only be inferred by soliciting its advice. The human can exploit by choosing to go ahead with an independent judgment. This choice is less risky when confidence in one's decision is high. However, when the AI's advice is not solicited, the human doesn't learn about the AI's ability. To appropriately judge relative expertise, it is necessary for the human to solicit the AI's help and make a mental model of the AI's ability.

COMPUTATIONAL MODEL: METACOGNITIVE BANDIT

The computational problem associated with the decision to solicit advice is an optimal exploration effort: Humans infer the relative utility of relying on themselves or the AI assistant to inform future decisions of when to seek help. This process be can be elegantly captured using the bandit framework. Bandit problems are widely used to study sequential decision-making when there is uncertainty about the rewards associated with decisions (or arms). In a machine learning context, multi-armed bandits have been used to efficiently choose between different sources of information, such as crowd workers and/or machine learning models

[1]Aakriti Kumar and Mark Steyvers are with the Department of Cognitive Sciences, University of California, Irvine, Irvine, CA 92697 USA, aakritk@uci.edu, msteyver@uci.edu

[2]Trisha Patel and Aaron Benjamin are with the Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL 61820 USA tpatel65@illinois.edu, asbenjam@illinois.edu

[8] and active assessment of machine classifiers [9]. In cognitive science, multi-armed bandits have been used to model human sequential decision behavior in reward and information seeking environments [10]–[12]. Decisions made in bandit problems require a balance between exploring all available arms and exploiting the best possible arm at any time.

We specify the decision to seek help from AI as a pull of one of two arms: self and AI. However, this decision is a metacognitive one: the human needs to evaluate their own performance (which will reflect the subjective difficulty of the current problem) as well as learn about the AI arm's utility. This is different from a traditional bandit setting in which the evaluation of arms corresponds to competing external events. The decision of arm selection on each trial is based on on the performance history of both arms (AI and self). The metacognitive bandit captures the metacognitive process employed by a human to decide whether to seek AI help on an individual trial. We hypothesize that the human infers a utility for soliciting the AI's help and a utility for coming up with a solution on their own. We use the framework of upper confidence bound (UCB) bandit models to model this process. Specifically, we use a Bayesian UCB framework [13] as a solution to this metacognitive task. In this framework, the decision-maker constructs a $100(1-\lambda)\%$ credible interval for the expected reward from each action at each trial and greedily chooses the action with the highest upper bound of the credible interval. It favors the exploration of actions with high uncertainty that have the potential to produce favorable outcomes. In our setting, at each trial t, the human compares the upper confidence bounds of relying on their own judgement or relying on the AI and pick the arm a with the higher inferred utility.

Let θ and ϕ denote the latent accuracy of the self arm (S) and the AI arm (AI) respectively. Let x_t denote the reward observed at each trial t for arm S and y_t denote the reward observed at each trial t for arm AI. The reward is 1 when an arm gives a correct response and 0 when the response is incorrect. Let a_t denote the action taken by the human where $a_t = 1$ if the AI was solicited on trial t and $a_t = 0$ if the AI was not solicited (i.e. the self arm was selected). We assume that the human always observes the reward for the self arm. However, for the AI arm, the correct and incorrect responses can only be observed for those trials when the arm was selected (i.e., a_t=1). This is an important feature of our model: the human always learns about their own ability but only learns about the AI's ability when the AI's help is solicited.

Our model further suggests that humans estimate a probability of being correct on the current stimulus without the AI's aid based on their inferred ability and the *subjective difficulty* of the stimulus at trial t. We use the term 'subjective difficulty' to draw attention to the possibility that an objectively easy trial can be perceived as a difficult trial by a human. This may happen because the human was not paying attention, or because the human doesn't have enough context or prior knowledge about a trial. The human infers a subjective difficulty for each stimulus presented. Let C_t be the true coherence level at time t. Perceived coherence ω_t is a sample from a normal distribution centered at C_t and standard deviation .2. We then impose an inverse transformation to estimate a subjective difficulty (d_t) based on the true coherence of a trial, $d_t = k/(\omega_t + \epsilon)$, where ϵ is a small value added to the denominator (set to .001 in our simulation) to avoid numerical issues. k is a proportionality constant set to .02. This equation gives us a way to estimate trial-level subjective difficulty for our experiment. The probability of being correct without the AI's help as estimated by the human is based on a Rasch model:

$$P(x_t = 1|\theta, d_t) = \frac{1}{1 + \exp(-(\theta - d_t))} \tag{1}$$

We use the sigmoid function to transform the value $(\theta - d_t)$ to a probability value between 0 and 1. This transformed density of the the latent ability serves as the distribution of expected accuracy for the self arm. In this model, the likelihood of observing a sequence of trial outcomes (i.e., runs of successes and failures) is:

$$p(X = x_{1:t-1}|\theta, d_{1:t-1}) = \prod_{j=1}^{t-1} \frac{\exp(x_j(\theta - d_j))}{1 + \exp(\theta - d_j)} \tag{2}$$

We assume that the human participant engages in an inference process about their own overall ability θ. Using Bayes's rule, the posterior over θ is:

$$p(\theta|X = x_{1:t-1}, d_{1:t-1}) \propto p(X_{1:t-1}|\theta, d_{1:t-1})p(\theta) \tag{3}$$

where we assume the prior $p(\theta) \sim N(\mu, \sigma^2)$. Since calculating the posterior exactly is intractable, we adopt an approximate inference technique to simulate the human's assessment of their own ability. We implement a Hamiltonian Monte Carlo algorithm to draw samples from the posterior of θ. The samples from the posterior are then used in equation 2 to infer the probability of being correct which adjusts for the difficulty of each particular trial.

As a simplifying assumption, we assume the human's inference about the AI's ability is independent of difficulty (as the human does not know what the AI finds difficult). The inference of the AI's ability is the same as the beta update in Equation 1.

To allow for some stochasticity in decision making, we assume that humans employ the Metacognitive bandit to choose the arm with the highest utility most frequently, but occasionally deviate from optimal behavior. The softmax action selection function is widely used to model uncertainty in human decision-making and gives us an elegant way to incorporate stochasticity in our model. After the adjustment for trial difficulty by the human, the probability of choosing the AI arm is evaluated using the softmax function:

$$p(a_t = 1) = \frac{1}{1 + \exp(-\frac{UCB(\sigma(\theta_t - d_t), \lambda) - (UCB(\phi_t, \lambda) - c)}{\tau})} \tag{4}$$

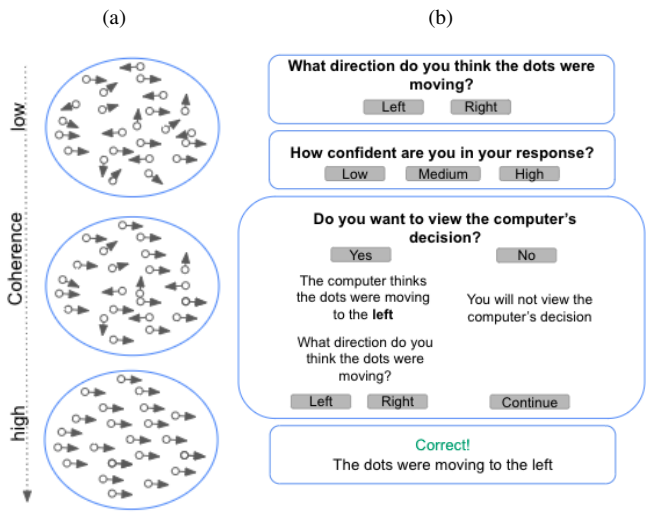

Fig. 1: Experimental setup: (a) Kinematograms with varying difficulties (coherence levels) were used as stimuli. (b) Sequence of events in the task.

where σ is the sigmoid function. In the model, the human observes a reward and updates the posterior of the latent ability θ, and ϕ for both arms. The subjective difficulty history $d_1, ..., d_{t-1}$, the current subjective difficulty d_t, and reward history $x_1, ..., x_{t-1}$ for arm S is observed at each time t. Similarly, reward history $y_1, ..., y_{t-1}$ is observed for arm AI (for those trials when the advice is solicited). Note that θ is conditioned on the history of the rewards $x_{1:t-1}$ accumulated by the human and the associated subjective difficulties $d_{1:t-1}$ of the trials, while ϕ is conditioned only on the history of the rewards $y_{1:t-1}$ accumulated by pulling the AI arm. We also impose a small cost $c = 0.1$ associated with the action of soliciting advice.

Illustrative Example

To illustrate the metacognitive bandit model's performance, we compare the predictions from the model to data collected from participants in a behavioral experiment. We provide a brief description of the sequence of events from one of a series of experiments on AI advice solicitation.

Participants were first shown a fixation point for 500 ms followed by a random-dot kinematogram for 500 ms (See Figure 1a). Participants were tasked with identifying the dominant direction of movement in the kinematogram (left or right). The coherence (randomness) of the kinematograms was randomly sampled from a uniform$(-.3, .3)$ distribution where negative coherence corresponds to left being the dominant movement direction of the stimuli. Low absolute value of coherence corresponds to higher trial difficulty. The sequence of events in the experiment as shown in Figure 1(b) were as follows. Participants were shown a kinematogram and were asked to submit an initial response. After submitting their response, they were asked to rate their confidence (low, medium or high) in their decision. Next, they were given the option to solicit the advice of an AI agent. If they chose to solicit advice, they were shown the AI recommendation. If not, they were shown feedback (correct/incorrect) on their original answer. If they solicited the AI's advice, they were allowed to change their answer after viewing the AI's recommendation. The AI advice did not include a confidence rating. AI advice was simulated by the experimenters such that AI accuracy increased as a function of coherence. Participants submitted their answer after taking into account the AI's advice. This was followed by feedback (correct/incorrect) on their final response. Note that a key feature of the experiment is that information about the AI's ability is only evident when its advice is solicited.

We focus on four qualitative findings related to participants' decision to solicit advice: First, human performance was on average poorer than that of the AI. Participants were correct 69% of the time on their first judgment, and the AI was correct 81% of the time. Therefore, participants are able to increase average performance by soliciting and adopting the advice of the AI. Note that both participants and AI did better on easier trials than harder trials. For example, for coherence values greater than .16, AI had an accuracy of 93% whereas humans had an accuracy of 78%. For coherence values less than .16, AI had an accuracy of 69% while humans had an accuracy of 61%. Second, there was substantial variation in the degree of soliciting AI advice across participants and trials. Figure 2(a) show that the tendency to solicit AI advice decreased over time. In addition, some individuals stopped soliciting advice after only a few trials, whereas other individuals kept soliciting advice across the entire experiment. Third, Figure 4(a) shows that confidence in the initial decision is related to accuracy, suggesting that participants have accurate metacognitive awareness of the difficulty of that particular trial and the associated level of uncertainty in their decision. Fourth, Figure 4(b) shows that AI advice was solicited more often when the participant was less confident. This is another way of saying that metacognition in this task is accurate — participants are able to judge the likelihood of answering correctly on a particular trial.

We simulated the metacognitive bandit by conditioning on the same true coherence and reward sequence as in the experimental data. We set λ to .1 such that the UCB action selection is based on the 90th quantile of the latent abilities of the two arms. We set c to .1 to impose a small cost associated with the action of soliciting advice. The experiment didn't ask participants for subjective difficulty on each trial. Instead, we use a noisy transformation of the true coherence of the stimuli used in the experiment to simulate subjective difficulty. Let C_t be the true coherence level at time t. Perceived coherence ω_t is a sample from a normal distribution centered at C_t and standard deviation .2. We then impose an inverse transformation to the perceived coherence to estimate a subjective difficulty of a trial, $d_t = k/(\omega_t + \epsilon)$, where ϵ is a small value added to the denominator (set to .001 in our simulation) to avoid numerical issues. k is a proportionality constant set to .02. This equation gives us

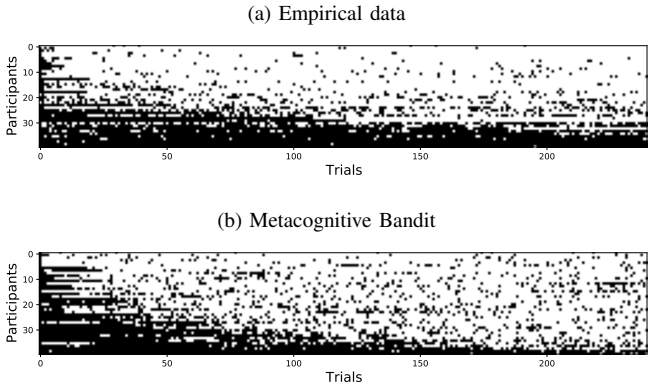

Fig. 2: Advice soliciting behavior for actual and simulated participants on 240 trials. Participants are sorted in increasing order of proportion of trials on which advice is solicited. White corresponds to trials where a participant did not solicit AI advice. (a) Empirical data (b) predictions of metacognitive bandit model

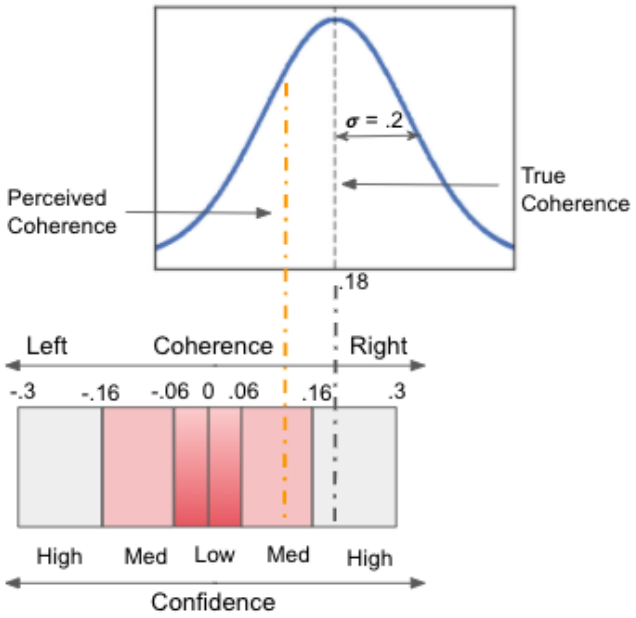

Fig. 3: Proposed generative model for human response and confidence: True coherence is sampled from a uniform distribution between $-.3$ and .3. Perceived coherence is a noisy sample from a normal centered at the true coherence and is used to determine the accuracy and confidence of the human on a trial.

a way to estimate trial-level subjective difficulty for our experiment. This is substituted in equation 1 to calculate the probability of being correct on each trial.

We also use the estimated perceived coherence of a trial to simulate the response and confidence of the human on that trial. Figure 3 shows the correspondence between the coherence value and the confidence of the human. We expect the human to have high confidence when the absolute value of coherence is high (between .16 and .3) and the direction of movement of the stimuli is highly discernible. We expect the human to have medium confidence when the absolute value of coherence is between .06 and .16 and low confidence when the absolute value of coherence is less than .6. If the human's perceived coherence has the same sign as the true coherence, we predict that the human can correctly guess the dominant direction of movement in the stimulus.

Through simulations, we see that the metacognitive bandit makes decisions similar to humans in the behavioral experiment, i.e, the model emulates the qualitative trends in the data. Figure 2(b) shows the advice seeking trend across the population simulated using the metacognitive bandit model. We see that the model predicts under reliance by some participants on AI advice. Figures 4 (a) and (b) also indicates that the model is able to capture the qualitative relationship between confidence, accuracy and the probability of seeking AI advice.

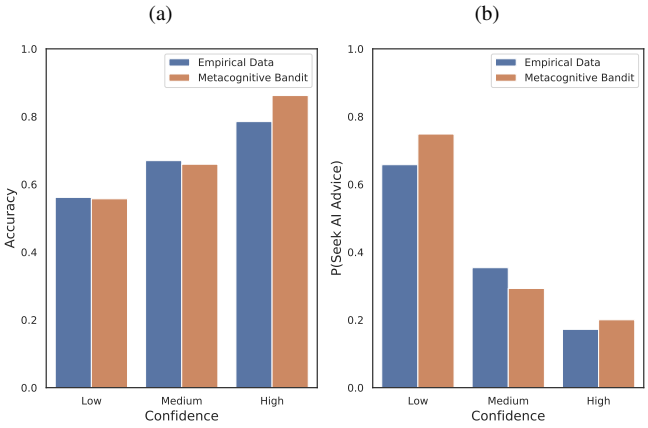

Fig. 4: Relationship between the reported confidence of participants in their response and (a) the accuracy of response, and (b) probability of soliciting AI advice.

Discussion

To build effective human-AI teams, in addition to using highly accurate and interpretable algorithms, it is critical to understand how humans how humans seek and use AI assistance. In this paper, we focus on understanding the cognitive process that drives a human's advice seeking behavior when working with AI. Through metacognitive bandits, we demonstrate that humans display a range of behaviors when working with an AI teammate. We show that individual differences in the reliance on the AI's advice can be expected when quasi-optimal decision-making strategies are applied to limited amounts of data observed from the AI. In the illustrative example discussed in this paper, we look at a very specific behavioral paradigm and use simulated AI advice. An important future direction is to look at more naturalistic decision-making settings while using a real AI in the loop. We also do not model how advice is integrated into the final decision by the human. Understanding how AI advice factors

into human judgment is another direction we plan to pursue.

Currently, our model only qualitatively captures trends in the data. To get a complete picture, we need to do more quantitative model fitting. However, we want to highlight the role that cognitive models can play in building more useful AI assistants. Cognitive models are tools to understand human intentions and knowledge. Modeling a human's understanding of an AI's ability can guide design of adaptive AI systems — it can be used to inform decisions about when and what kind of assistance should be provided to the user. Understanding how humans acquire machine theory of mind and using that knowledge to endow AI with a model of the user is an important step towards building better human-AI teams. Ultimately, models of human-AI interaction will be critical for understanding human behavior in hybrid teams and also for designing AI agents in a way that humans can use most effectively.

REFERENCES

[1] E. Kamar, "Directions in hybrid intelligence: Complementing ai systems with human intelligence." *IJCAI*, pp. pp. 4070–4073, 2016.

[2] B. Green and Y. Chen, "Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 2019, pp. 90–99.

[3] S. Bonaccio and R. S. Dalal, "Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences," *Organizational behavior and human decision processes*, vol. 101, no. 2, pp. 127–151, 2006.

[4] F. Gino and D. A. Moore, "Effects of task difficulty on use of advice," *Journal of Behavioral Decision Making*, vol. 20, no. 1, pp. 21–35, 2007.

[5] J. M. Logg, "Theory of machine: When do people rely on algorithms?" *Harvard Business School working paper series# 17-086*, 2017.

[6] J. M. Logg, J. A. Minson, and D. A. Moore, "Algorithm appreciation: People prefer algorithmic to human judgment," *Organizational Behavior and Human Decision Processes*, vol. 151, pp. 90–103, 2019.

[7] B. J. Dietvorst, J. P. Simmons, and C. Massey, "Algorithm aversion: People erroneously avoid algorithms after seeing them err." *Journal of Experimental Psychology: General*, vol. 144, no. 1, p. 114, 2015.

[8] L. Tran-Thanh, S. Stein, A. Rogers, and N. R. Jennings, "Efficient crowdsourcing of unknown experts using bounded multi-armed bandits," *Artificial Intelligence*, vol. 214, pp. 89–111, 2014.

[9] D. Ji, R. L. Logan IV, P. Smyth, and M. Steyvers, "Active bayesian assessment for black-box classifiers," in *35th AAAI Conference on Artificial Intelligence*, 2021.

[10] M. Steyvers, M. D. Lee, and E.-J. Wagenmakers, "A bayesian analysis of human decision-making on bandit problems," *Journal of Mathematical Psychology*, vol. 53, no. 3, pp. 168–179, 2009.

[11] M. Speekenbrink and E. Konstantinidis, "Uncertainty and exploration in a restless bandit problem," *Topics in cognitive science*, vol. 7, no. 2, pp. 351–367, 2015.

[12] C. M. Wu, E. Schulz, M. Speekenbrink, J. D. Nelson, and B. Meder, "Generalization guides human exploration in vast decision spaces," *Nature human behaviour*, vol. 2, no. 12, pp. 915–924, 2018.

[13] N. G. Pavlidis, D. K. Tasoulis, and D. J. Hand, "Simulation studies of multi-armed bandits with covariates," in *Tenth International Conference on Computer Modeling and Simulation (uksim 2008)*. IEEE, 2008, pp. 493–498.