Leveraging Cognitive Models for the Wisdom of Crowds in Sequential Decision Tasks

Erin H. Bugbee (ebugbee@cmu.edu)
Chase McDonald (chasemcd@cmu.edu)
Cleotilde Gonzalez (coty@cmu.edu)
Department of Social and Decision Sciences, Carnegie Mellon University
Pittsburgh, PA 15213 USA

Abstract

Many decisions we face in life are sequential, where alternatives appear over time. We often must decide whether to take the opportunity and stop searching or to continue evaluating potentially better future alternatives. Humans are notoriously poor at stopping optimally in sequential decision-making tasks. These sequential decisions are difficult because they involve the consideration of how past, present, and future decisions affect the outcome. Recent research suggests that the wisdom of the crowd (WoC) — that is, aggregated decisions of many people that outperform most individuals — can be applied to sequential decision tasks and potentially help improve stopping decisions. Current models rely on a process of fitting human data, making it difficult to understand how those individuals would behave in new problems. Furthermore, these models do not account for the learning process that humans experience while making these decisions. In this work, we demonstrate how simulated agents using a cognitive model derived from Instance-Based Learning Theory (IBLT) can produce WoC that is similar to WoC from human participants in two sequential decision tasks. We demonstrate that the WoC performance from simulated groups of agents is better than the performance of most agents and that the Instance-Based Learning (IBL) crowd behavior is similar to the human crowd behavior. Thus, cognitive models that account for learning and experience can be used to inductively predict the behavior of human crowds in sequential decision tasks.

Keywords: wisdom of crowds; sequential decision making; cognitive modeling; instance-based learning

Introduction

Sequential decision making is ubiquitous in everyday life. As we navigate the world and make decisions, we often do not face all possible alternatives at once. Instead, alternatives emerge over time, and to maximize benefits, a choice must involve the selection of an alternative at the right time, before the opportunity disappears. For example, to select a rental apartment in a dynamic market, one must decide when to stop visiting new possibilities and make an offer before the current option becomes unavailable.

The literature on sequential decisions has underscored that people are often suboptimal in making stopping decisions in sequential tasks, given the tradeoffs of risk and uncertainty (Lejuez et al., 2002; Lee, 2006; Guan, Stokes, Vandekerckhove, & Lee, 2020; Guan, 2019; Bugbee & Gonzalez, 2022b). Recently, the possibility of using the Wisdom of Crowds (WoC) has been suggested as a way to address these difficulties in sequential decision problems (Thomas, Coon, Westfall, & Lee, 2021). The WoC (Surowiecki, 2005) suggests that the aggregation of individual estimates or decisions can outperform most of the individuals in the crowd, and a significant amount of work has demonstrated that the aggregation of collective wisdom can be beneficial in a large number of tasks. However, the benefits of WoC for sequential decision tasks have only recently been suggested (Thomas et al., 2021). The idea is to aggregate the answers from a group of individuals in each choice of a sequence to produce a crowd answer (e.g., whether to stop exploring or not), and such aggregate would produce an answer closer to the optimal stopping point compared to individual decisions.

In their work, Thomas et al. (2021) aggregate individual predictions to retrieve WoC predictions. These WoC predictions, along with the individual responses, are then compare to the predictions of cognitive models at the individual and crowd level. They used statistical models of individuals to provide model-based predictions, showing that the aggregation of these predictions can result in accurate behavior in not only problems that individuals completed, but also new problems that participants did not previously experience. The models that Thomas et al. (2021) present are descriptive statistical models, where the parameters of the models are fit to the data of individuals, and these parameter values are then used to generalize to new problems within the same class of sequential decision-making problems. These are not process models that represent the individual learning through a sequence; and thus, they would fail to account for behavior in situations in which people learn from past choices.

In this research, we build on the work of Thomas et al. (2021) to test the benefits of WoC using cognitive models. We rely on two known sequential decision tasks and their previously collected data sets (Guan, 2019; Guan et al., 2020). Further, we utilize two existing cognitive models of sequential decisions in these tasks (Bugbee & Gonzalez, 2022a, 2022b). In contrast to the work of Thomas et al. (2021), these cognitive models are generative process models that learn through experience to produce predictions of human stopping decisions in the absence of human data. These models act based on a theory of decisions from experience, *Instance-Based Learning (IBL) Theory* (Gonzalez, Lerch, & Lebiere, 2003), and are able to replicate human sequential decisions closely. The question in this research is whether the WoC predictions in groups of agents generated with IBL models result in similar values as the groups of human participants in the same sequential decision making tasks. The replication of the human crowd behavior in addition to the individual human behavior has significant benefits for applying the WoC to sequences of decisions in new situations for which human data might not exist.

Sequential Decision Tasks and Data Sets

For this work, we used experimental data previously collected by Guan (2019) and Guan et al. (2020) in two sequential decision tasks: the Balloon Analog Risk Task (BART) and the Optimal Stopping Task.

Balloon Analog Risk Task (BART)

BART is a sequential decision making task in which a decision maker inflates a balloon. The level of inflation corresponds to the reward that the decision maker can receive. At each time point, the decision maker decides whether to pump the balloon and increase its value or bank the current monetary amount. However, with each pump of the balloon, there is a probability that the balloon bursts, causing the decision maker to receive a reward of 0 for that problem. This leads to the need to balance exploring through pumping with exploitation through banking, with the goal of maximizing total reward. Each balloon has a predefined burst time generated from the constant probability of bursting, although participants are not told these probabilities.

In the experiment from Guan et al. (2020), 56 participants completed the BART in a within-subjects design. Participants were presented balloons with a fixed probability of bursting with each pump (either P(Burst) = 0.1 or P(Burst) = 0.2)[1]. Every participant completed 50 problems with each probability, and the order of the problems and conditions was randomized between participants. Each problem started with a balloon with a hypothetical value of $1. For each decision, the participant had the option to pump the balloon ("Pump") and increase its monetary value by $1, or stop ("Bank") and collect the current monetary value. However, each pump action risks bursting the balloon, which results in collecting $0 for that problem. The participant continued making Bank or Pump decisions until either the balloon burst or the participant chose the Bank action and collected the money. The stated goal was to maximize the total reward on all problems. Participants were compensated for their time but were not rewarded based on their performance.

Optimal Stopping Task

In the Optimal Stopping Task, the same 56 participants from Guan et al. (2020) were presented with sequences of cats. They were instructed to choose the cat in the sequence with the highest weight. Participants were presented with cats sequentially and were required to "Select" or "Pass' each cat. Once passed, the cat could not be returned to. The participants were instructed that the last cat in the sequence must be chosen if none is chosen prior. If they chose the cat with the maximum weight, they were correct, and otherwise they were incorrect. Participants received feedback on the accuracy of their choice, but not about unseen cats.

[1]This deviates from the typical BART design (e.g. Lejuez et al. (2002)), in which the probability of the balloon bursting increases as the number of pumps increases.

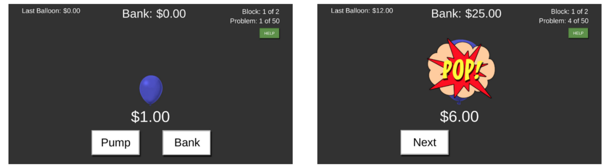


Figure 1: A screenshot of the BART experiment obtained from Guan (2019).

The task had a within-subjects design. Each participant experienced four conditions, in which both the distribution of weights and the length of the sequence were varied. Weights ranged between 0 and 100 pounds, according to either a uniform (i.e., "neutral") or beta(4,2) (i.e., "plentiful", as weights are skewed toward higher weights) distribution scaled to 0 to 100. The sequences consisted of 4 or 8 cats. Participants were told the length of the sequence but not the distribution of cat weights.

All participants completed a group of 40 problems within each condition with a randomized problem order among the participants. The order of conditions was also randomized among participants.

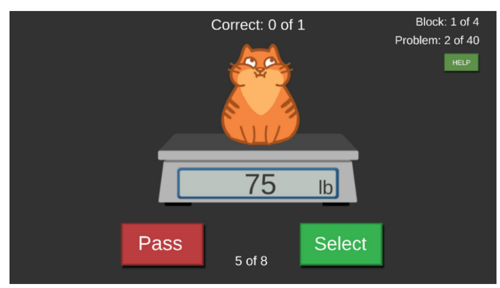


Figure 2: A screenshot of the Optimal Stopping Task experiment obtained from Guan (2019).

Instance-Based Learning Theory

We use cognitive models based on Instance-Based Learning Theory (IBLT) recently implemented in Bugbee and Gonzalez (2022b) and Bugbee and Gonzalez (2022a). IBLT outlines a theory of decisions from experience, derived from mechanisms proposed in the ACT-R cognitive architecture (Anderson & Lebiere, 2014). The theory was developed to explain human learning in dynamic environments (Gonzalez et al., 2003). It provides an algorithm for learning from experience and making decisions, which can be used to implement a computational model of these processes that simulates human behavior.

There are three primary components of the decision making algorithm: recognition and retrieval of past instances, as a function of their similarity to a current decision; calculation of the expected utility of decision alternatives, and a choice rule that allows for generalization from past experience. Past instances are stored in memory and are effectively memory units consisting of situations $s \in S$, decisions $a \in A$, and the realized utility x of taking action a after observing situation s. An option is defined as $k = (S,A)$: making a decision A in the situation S.

At time t, there are $n_{k,t}$ different generated instances $(k, x_{i,k,t})$ for $i = 1,...,n_{k,t}$, corresponding to selecting k and achieving the outcome $x_{i,k,t}$. Each instance i in memory has an activation value, which represents how readily available this information is in memory, and is determined by similarity to past situations, recency, frequency, and noise (Anderson & Lebiere, 2014). The activation is described by Equation 1, for option j, when presented with option k (that is, the current situation is described by k):

$$\Lambda_{i,k,j,t} = \ln\left(\sum_{t' \in T_i} (t-t')^{-d}\right) + \alpha S(k,j) + \sigma \ln \frac{1-\xi}{\xi} \quad (1)$$

where

$$S(k,j) = \sum_j Sim_j(f_j^k, f_j^j) \quad (2)$$

and α, d and σ are the mismatch penalty, decay, and noise parameters, respectively. Furthermore, $T_i \subset \{0,...,t-1\}$ is the set of previous timestamps in which the instance i was observed and Sim_j is a similarity function that calculates the similarity of the jth attribute of an option k, f_j^k. The rightmost term represents Gaussian noise to capture individual variation in activation, and ξ is a random number drawn from a uniform distribution $U(0,1)$ at each time step and for each instance and option.

The probability of retrieving an instance i from memory is a function of its activation $\Lambda_{i,k,j,t}$ relative to the activation of all instances:

$$p_{i,k,j,t} = \frac{\exp(\frac{\Lambda_{i,k,j,t}}{\tau})}{\sum_{j=1}^{n_{j,t}} \exp(\frac{\Lambda_{j,k,j,t}}{\tau})} \quad (3)$$

where τ is the temperature parameter. As $\tau \to 0$, the selection of actions is deterministic, and as $\tau \to \infty$, all actions become equally likely.

The expected utility of option k is given by the *blending* mechanism calculated as in Gonzalez and Dutt (2011):

$$V_{k,t} = \sum_{i=1}^{n_{j,t}} p_{i,k,j,t} x_{i,j,t}. \quad (4)$$

The blending operation (Equation 4) is the sum of all past experiences weighted by their probability of retrieval, for which the option with the maximum blended value is selected greedily. In particular, at the l-th step of an episode, the agent selects the option (s_l, a_l) with

$$a_l = \arg\max_{a \in A} V_{(s_l,a),t} \quad (5)$$

When the agent receives delayed results, the agent updates expected utilities using a credit assignment mechanism (Nguyen, McDonald, & Gonzalez, 2021). Throughout the present work, we use default parameter values for decay $d = 0.5$ and noise $\sigma = 0.25$. The mismatch penalty α is set for each task individually.

IBL Model of BART

We use a previously developed IBL model for the BART (Bugbee & Gonzalez, 2022b). The instance structure in this model is as follows: the situation has the feature of the number of pumps of the balloon prior to the present decision, the decision is to pump the balloon or bank, and the utility depends on the outcome of that decision. If the balloon bursts from that decision, then the utility is 0, since the model should learn that pumping at that number of pumps led to bursting the balloon and receiving no money for that problem. If the balloon does not burst from that decision, then the utility is the value of the balloon or the number of pumps thus far plus one for the initial value, since the model should learn that pumping at that number of pumps did not burst the balloon. The model uses partial matching, in particular linear similarity, to compare the current instance to past ones, and a mismatch penalty of $\alpha = 5$.

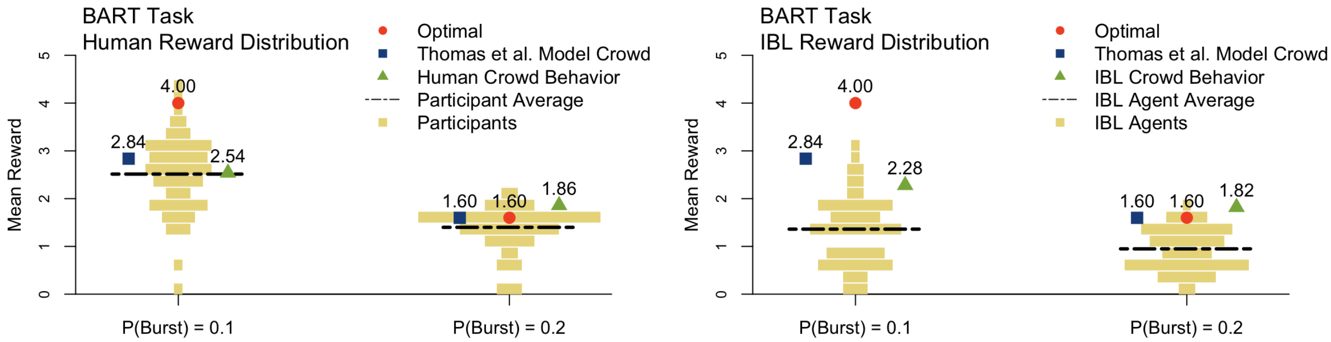
IBL Model of Optimal Stopping Task

We similarly use an IBL model for the Optimal Stopping Task proposed by Bugbee and Gonzalez (2022a). The instance structure of this model is as follows: the situation has the feature of the value of the current alternative and the number of alternatives remaining in the sequence, the decision is to select the alternative or pass, and the utility is 1 if the selected alternative is the maximum and 0 otherwise. The model uses a credit assignment mechanism such that the utility is propagated back to the previous decisions in the sequence once a select action is made and the outcome is observed. The model uses partial matching, in particular linear similarity, to compare the current instance to past instances, and a mismatch penalty of $\alpha = 10$.

Model Simulation Methods

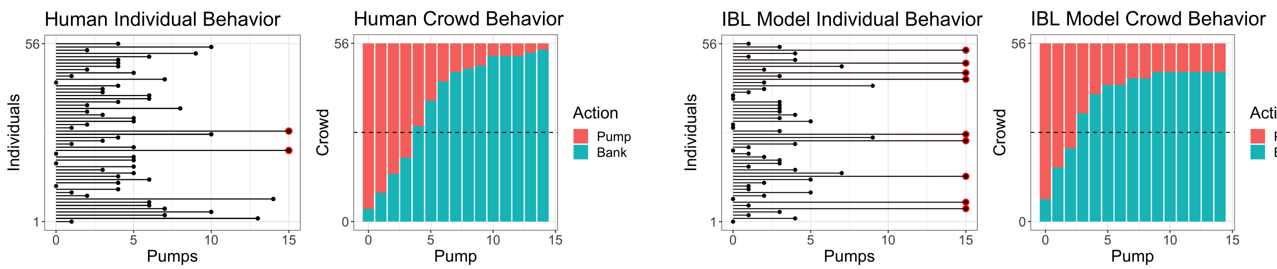
We use cognitive models based on IBLT (Gonzalez et al., 2003) and implemented using PyIBL, a Python implementation of IBLT (Morrison & Gonzalez, 2021). As mentioned, these models were developed and reported in Bugbee and Gonzalez (2022b) for the BART and Bugbee and Gonzalez (2022a) for the Optimal Stopping Task.

For each human participant in the data set, we simulate an IBL model agent experiencing the same stimuli, that is, the exact problems and conditions in the same order as the human. Therefore, we can map each IBL model agent to a corresponding human participant in the original study. Importantly, the models are not fit to the human data, so the correspondence between models and humans is only a result of the similarity of their experiences. As a result, we have 56 simulated IBL model agents making choices in each task,



(a) (b)

Figure 3: (a) Distribution of mean rewards for the human participants and (b) distribution of mean rewards for IBL model agents in the BART task, compared to the Thomas et al. (2021) Model Crowd and the optimal decision process.



(a) (b)

Figure 4: (a) Individual behavior (left panel) and crowd behavior (right panel) for the human participants and (b) individual and crowd behavior for IBL agents for a problem in the BART task in which the balloon bursts at pump 15. For the individual behavior, lines indicate the number of pumps for each individual, and red dots indicate that the participant burst the balloon. For the crowd behavior, the dashed line corresponds to half of the participants to visualize the majority decision. The human crowd banks at pump 4 and the IBL model crowd banks at pump 3.

where each simulated agent maps directly to a particular human participant.

Wisdom of Crowds Aggregation

In alignment with Thomas et al. (2021) we use the behavior-based majority decision to determine the WoC decision. That is, for each decision, the behavior of the crowd is that of the majority of participants.

In the BART, the behavior-based WoC crowd behavior is governed by the majority decision to pump or bank on each trial. Each individual, given that they have not already banked, decides whether to pump the balloon or bank the money. Once a participant decides to bank, presumably they have decided to bank on all following decisions, so we impute those after banking as bank decisions as well. This is a deviation from Thomas et al. (2021), where they remove participants after they make a bank decision. The crowd follows the majority until the majority either banks or the balloon bursts.

In the Optimal Stopping Task, the behavior-based WoC behavior depends on the majority decision at each alternative. For a particular alternative, each individual decides to select that alternative or pass and see the next one. The crowd will select the alternative if that is the selection of the majority; otherwise, it will pass and continue until either the majority selects a particular cat or the end of the sequence is reached and the last cat must be chosen. This is directly in alignment with Thomas et al. (2021).

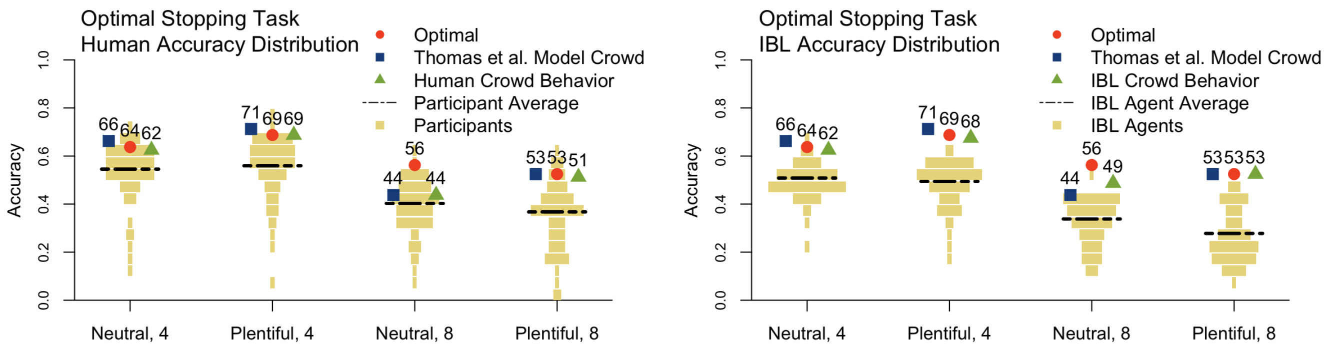
Results

For the results, we will show the individual behavior for the human participants and IBL agents alongside their respective crowd behaviors corresponding to the majority decisions. For the BART, the crowd decision is determined for each pump or bank decision. For the Optimal Stopping Task, the crowd decision is determined at each select or pass decision.

WoC in the BART Task

Figures 3a and 3b show the distribution of mean rewards, the average reward, and the crowd behavior for the human participants and IBL agents respectively. The figures also display the optimal reward and the reward of the model crowd from Thomas et al. (2021) for comparison.

The distribution of mean rewards is slightly lower for the IBL model than for humans. This is explained by the need for the IBL model to learn from experience how to gain points without “reading instructions” while human participants read



(a) (b)

Figure 5: (a) Distribution of mean rewards for human participants and (b) distribution of mean rewards for IBL model agents in the Optimal Stopping Task, compared to the Thomas et al. (2021) Model Crowd and the optimal decision process.

instructions and therefore start with more understanding of the task. The crowd behavior is comparable for the human and IBL model in both conditions. We see that the crowd performs better than the average across participants in both conditions for the human and IBL models. These results indicate that the crowd performs this task better than the average individual, and the IBL crowd behavior closely replicates the human crowd.

The "Thomas et al. Model Crowd" represented by the blue square in the figures comes from Thomas et al. (2021), and it is based on the Two-Parameter BART model (van Ravenzwaaij, Dutilh, & Wagenmakers, 2011). This model assumes that participants have a target number of pumps for each problem that they do not adapt over problems, which depends on their risk propensity and belief about the burst probability of the balloon (for more details, see Thomas et al. (2021)). Ultimately, the IBL crowd behavior shows improved performance over that of the Two-Parameter BART model in the P(Burst) = 0.2 condition, and has slightly worse performance in the P(Burst) = 0.1 condition.

The optimal performance represented by the red circle was determined by Monte Carlo simulation in Thomas et al. (2021), as the optimal number of pumps is challenging to derive. This shows 10 pumps to be optimal for P(Burst) = 0.1 yielding around \$4.00 on average, and 4 pumps to be optimal for P(Burst) = 0.2, yielding around \$1.60. Thomas et al. (2021) explained that the optimal performance appears low because the problems used in Guan et al. (2020) are fairly unrepresentative of the true environment. As many problems had late burst trials, it is possible to perform better than optimal, which we see for some human individuals and the human crowd, as well as for some IBL agents and the IBL crowd.

Figure 4 shows an example of the behavior of humans (a) and IBL models (b) in which the balloon bursts at pump 15. We observe comparable pumping behavior for humans and IBL agents. In this problem we see more IBL agents pumping more (up to pump 15) when the balloon bursts. But we also observe that for the crowd behavior the human crowd banks at pump 4 and the IBL model crowd banks at pump 3.

Optimal Stopping Task

Figures 5a and 5b show the accuracy distribution for humans and IBL model agents in the four conditions of the optimal stopping task. We observe similar distributions of accuracy between participants and IBL model agents.

We also see that crowd behavior in the IBL model is comparable to that of the human participants — in fact, the IBL model crowd performance is better in the Length 8 conditions relative to the performance of the humans. The crowd behavior is better than the average participant in all conditions for both the human and the IBL model agents. This indicates that the WoC is better than the average participant, and that the IBL WoC closely replicates the WoC of human participants.

The "Thomas et al. Model Crowd" from Thomas et al. (2021), represented by the blue square, is based on three fixed-then-linear strategies used to set thresholds for making stopping decisions. That is, participants may have fixed thresholds over positions and choose the first alternative that exceeds that threshold; they may have a starting threshold which they decrease linearly throughout the sequence; or they may have a fixed threshold for some fixed trials in the sequence, which they then decrease linearly. It is assumed that a participant uses the same strategy for all problems. The relationship between the human WoC and the Thomas et al. Model Crowd is similar to that of the IBL model WoC and the Model Crowd, again suggesting that the IBL model can replicate the human crowd behavior.

The optimal performance represented by the red circle was determined according to the findings of Gilbert and Mosteller (1966), as reported in Thomas et al. (2021). The optimal strategy is to choose the first value that is the current maximum in the sequence and is above the optimal threshold calculated based on the position in the sequence. Thomas et al. (2021) clarify that the optimal performance is surpassed since there are a finite number of experimental problems. We see that both individuals and the various crowds sometimes have comparable or even greater accuracy than the optimal strategy.

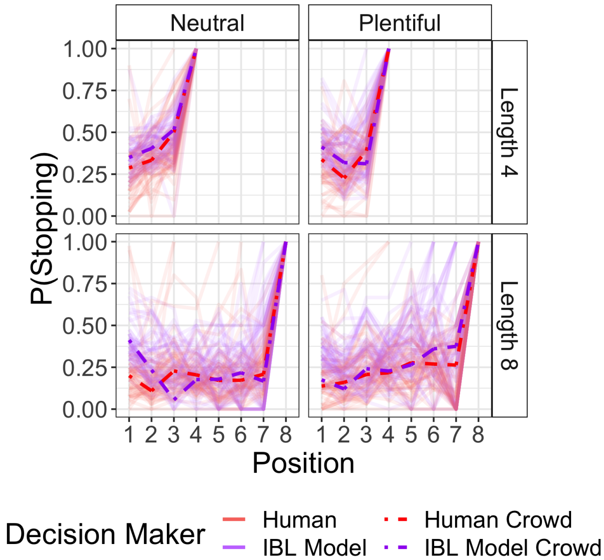


Figure 6: Probability of stopping by position, for individuals and the crowd for the human participants and IBL agents.

Figure 6 shows the stopping probabilities in the Optimal Stopping Task at each position in the sequence for the four conditions. Each solid line corresponds to a particular individual's stopping probabilities over all problems. The dashed lines correspond to the stopping probabilities of the crowd. The similarity of the IBL agents to the human participants, as well as the IBL crowd and the human crowd, indicates that the IBL model is able to capture the stopping probability of human participants under each condition, and that the IBL model is able to replicate the human crowd closely.

Discussion

The WoC, involving an aggregation of individual decisions, has been shown to be a powerful and effective method of producing results that are better than many people in that group (Surowiecki, 2005). However, it is unclear whether this wisdom can be beneficial in sequential decision-making tasks. Recent research suggests that simple aggregation rules (e.g., a majority decision) can result in more optimal stopping decisions in these tasks compared to the stopping decisions of most individuals in the group (Thomas et al., 2021).

This research builds on the work of Thomas et al. (2021) by demonstrating that it is possible to use cognitive models to simulate a crowd of agents and that the WoC resulting from the simulated crowd is similar to the WoC resulting from human participants. We demonstrate this idea in two sequential decision tasks. Importantly, in contrast to the descriptive statistical models in Thomas et al. (2021), we employ the learning models based on a cognitive theory of decisions from experience, IBLT (Bugbee & Gonzalez, 2022a, 2022b). The simulation results demonstrate how these models provide predictions of human behavior and that the WoC derived from the aggregation of the simulated agents results in improved performance relative to the individual agents. Importantly, the WoC predictions of the model are similar to the WoC calculated from human data.

The cognitive models we utilize for WoC are learning models, and this addresses a primary limitation described by Thomas et al. (2021), in that their statistical models could not dynamically adapt as human decision makers. Although Thomas et al. (2021) show that models that fit human data can generalize to problems in the same class of tasks, our work demonstrates that a cognitive model that accounts for learning without relying on specific human data can be used across distinct tasks of varying structure, while providing comparable individual-level predictions and WoC decisions.

Learning is likely to occur in human participants to some extent, and there is value in being able to capture behavioral changes as their experience grows. Our results demonstrate that the cognitive models we propose can learn to perform at the same level as human participants and that the WoC derived from crowds of IBL agents are similar to the WoC derived from human crowds. In future work, these models could be applied to settings in which human adaptation is a prominent feature of the task.

Acknowledgments

An Open Science Framework project is available at `https://osf.io/275gp/` with the data, code, and analysis files. This research was supported by AFRL Award FA8650-20-F-6212, sub-award number 1990692 to Cleotilde Gonzalez. The authors thank John Anderson, Dan Bothell, Christian Lebiere and Leslie Blaha for their helpful feedback regarding this work. We also thank Michael Lee for assistance with the data.

References

Anderson, J. R., & Lebiere, C. J. (2014). *The atomic components of thought*. Psychology Press.

Bugbee, E. H., & Gonzalez, C. (2022a). Deciding when to stop: Cognitive models of sequential decisions in optimal stopping tasks. *In preparation*.

Bugbee, E. H., & Gonzalez, C. (2022b). Making predictions without data: How an instance-based learning model predicts sequential decisions in the balloon analog risk task. In *Proceedings of the annual meeting of the cognitive science society*.

Gilbert, J. P., & Mosteller, F. (1966). Recognizing the maximum of a sequence. *Journal of the American Statistical Association*, *61*(313), 35–73. Retrieved from `http://www.jstor.org/stable/2283044`

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating decisions from experience in sampling and repeated choice paradigms. *Psychological Review*, *118*(4), 523–51.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*. doi: 10.1016/S0364-0213(03)00031-4

Guan, M. (2019). *A cognitive modeling analysis of risk in sequential choice tasks.* Retrieved from `https://escholarship.org/uc/item/802684nb`

Guan, M., Stokes, R., Vandekerckhove, J., & Lee, M. D. (2020). A cognitive modeling analysis of risk in sequential choice tasks. *Judgment and Decision Making*, *15*, 823-850.

Lee, M. D. (2006). A hierarchical bayesian model of human decision-making on an optimal stopping problem. *Cognitive science*, *30*(3), 1–26.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., ... Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the balloon analogue risk task (bart). *Journal of Experimental Psychology: Applied*, *8*(2), 75.

Morrison, D., & Gonzalez, C. (2021). *Pyibl version 4.0.* Retrieved 2021-03-30, from `http://pyibl.ddmlab.com/`

Nguyen, T. N., McDonald, C., & Gonzalez, C. (2021). *Credit assignment: Challenges and opportunities in developing human-like ai agents* (Tech. Rep.). Carnegie Mellon University.

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.

Thomas, B., Coon, J., Westfall, H., & Lee, M. (2021, 07). Model-based wisdom of the crowd for sequential decision-making tasks. *Cognitive Science*, *45*, e13011. doi: 10.1111/cogs.13011

van Ravenzwaaij, D., Dutilh, G., & Wagenmakers, E.-J. (2011). Cognitive model decomposition of the bart: Assessment and application. *Journal of Mathematical Psychology*, *55*(1), 94-105. (Special Issue on Hierarchical Bayesian Models) doi: https://doi.org/10.1016/j.jmp.2010.08.010