Task Allocation in Teams as a Multi-Armed Bandit

Raja Marjieh*
raja.marjieh@princeton.edu
Princeton University
Princeton, NJ, USA

Anand Gokhale*
anand_gokhale@ucsb.edu
UC Santa Barbara
Santa Barbara, CA, USA

Francesco Bullo
bullo@ucsb.edu
UC Santa Barbara
Santa Barbara, CA, USA

Thomas L. Griffiths
tomg@princeton.edu
Princeton University
Princeton, NJ, USA

ABSTRACT

Humans rely on efficient distribution of resources to transcend the abilities of individuals. Successful task allocation, whether in small teams or across large institutions, depends on individuals' ability to discern their own and others' strengths and weaknesses, and to optimally act on them. This dependence creates a tension between exploring the capabilities of others and exploiting the knowledge acquired so far, which can be challenging. How do people navigate this tension? To address this question, we propose a novel task allocation paradigm in which a human agent is asked to repeatedly allocate tasks in three distinct classes (categorizing a blurry image, detecting a noisy voice command, and solving an anagram) between themselves and two other (bot) team members to maximize team performance. We show that this problem can be recast as a combinatorial multi-armed bandit which allows us to compare people's performance against two well-known strategies, Thompson Sampling and Upper Confidence Bound (UCB). We find that humans are able to successfully integrate information about the capabilities of different team members to infer optimal allocations, and in some cases perform on par with these optimal strategies. Our approach opens up new avenues for studying the mechanisms underlying collective cooperation in teams.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative and social computing theory, concepts and paradigms**; • **Applied computing** → **Psychology**.

KEYWORDS

Task Allocation, Team Coordination, Multi-Armed Bandit

ACM Reference Format:
Raja Marjieh, Anand Gokhale, Francesco Bullo, and Thomas L. Griffiths. 2024. Task Allocation in Teams as a Multi-Armed Bandit. In *Proceedings of Collective Intelligence (CI '24).* ACM, New York, NY, USA, 3 pages.

*Both authors contributed equally to this research.

CI '24, June 26–29, 2024, Boston, MA

1 TASK ALLOCATION PARADIGM

Our paradigm is inspired by the literature on Transactive Memory Systems (TMS) [9]. In a typical TMS setup [4], a team of individuals with initially unknown skill levels attempt to solve a sequence of tasks and gradually form a collective representation of the expertise of team-members. Here we adapted this process to study learning in a single individual and to allow for theoretical tractability (though see Discussion). To do so, participants were instructed to allocate tasks from three different classes between them and two other team members across 20 iterations, such that in a given iteration each team member receives a single task from a distinct class (the participant decides on the class allocation and then a random task is sampled; Figure 1A).

To decouple exploration from memory, team performance was summarized in a status board and the members with highest empirical task success rates were highlighted. The puzzle classes covered three modalities: visual, auditory, and lexical. In the visual task, the agent had to categorize an object from a blurry image. Here the images were sampled from CIFAR-10H [5] which contained natural images from the categories: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. To ensure that the task is challenging, we selected images with medium classification entropy based on prior human classifications in CIFAR-10H. In the auditory task, participants had to detect an audio command in a noisy signal. The commands were taken from the Speech Commands Dataset [8] and comprised the words: up, down, go, stop, left, right, yes, no. To make the task harder we added white noise to the signals at -12dB SNR. Finally, for the lexical task, participants had to solve an anagram. Anagrams were created by selecting 211 common 5-letter words and shuffling their letters.

The other team members were modeled as bots with 70% chance of succeeding for one class of tasks and 15% at the others to ensure that there is a single optimal allocation. We ran three experimental batches, one for each allocation of bot skill levels to task classes. Moreover, participants received a performance bonus in proportion to their team score. Overall, we recruited 300 participants from Prolific, and they all provided informed consent prior to participation in accordance with an approved Princeton University Institutional Review Board (IRB) protocol (#10859).

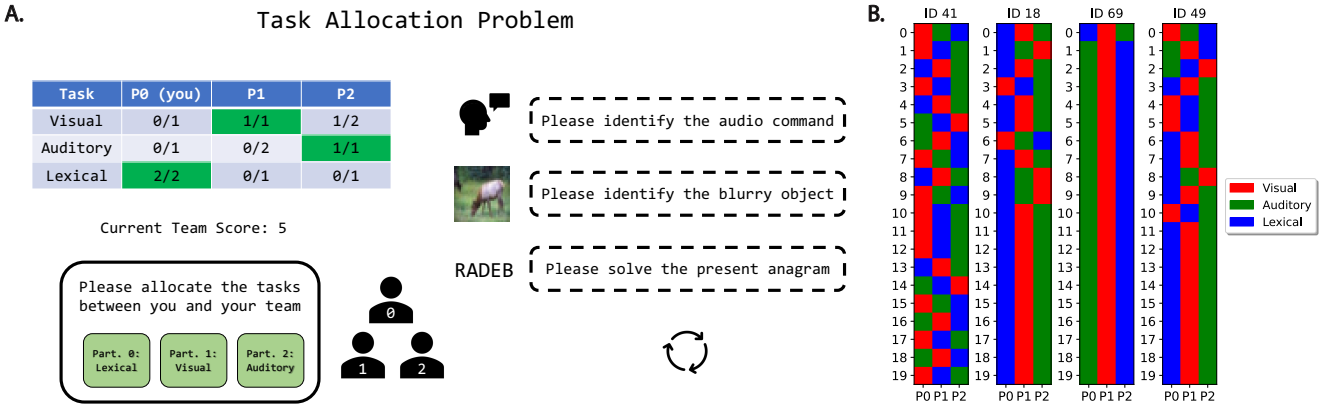

Figure 1: Task Allocation Paradigm. (A) Schematic of the task. (B) Example human allocation dynamics.

2 THEORETICAL FORMULATION

Our paradigm presents participants with a multi-armed bandit (MAB) problem with interdependent arms due to shared agent-task pairs (i.e., each arm corresponds to an allocation of all team members to task classes, and some allocations overlap). Classic optimal approaches to the MAB problem such as UCB [1] or Thompson sampling [3] are designed for *independent* arms. Further, participants receive feedback on each team member's performance. Our paradigm, therefore, is better described by a combinatorial semi-bandit [7] where each unique agent-task pair corresponds to an arm, and the participant must choose among valid combinations of those arms (i.e., an allocation where each member receives one unique puzzle class), which may be thought of as *superarms*. Depending on whether (combinatorial) UCB or Thompson sampling is used, at each timestep, each superarm $\mathcal{S}$ is associated with a score

$$\text{Score}(\mathcal{S}) = \sum_{x \in S} U_A(\mu(x), \sigma(x)) \tag{1}$$

where $\mu(x)$ is the estimate for the mean score for the agent-task pair x, $\sigma(x)$ is the uncertainty in our estimate, and U_A is the value function for algorithm A (UCB or Thompson) and is calculated as if the arm were independent [7]. The superarm with the maximum score is selected at each timestep. We note that modeling of human decision-making in MAB problems is well-studied in simple bandits [2] as well as contextual bandits [6]. However, to our knowledge, combinatorial bandits have not previously been used to study task assignment problems.

3 RESULTS

Our paradigm yielded rich exploration dynamics despite its simplicity. Figure 1B shows data from four prototypical individuals. We see that different participants varied in their approach to the problem. Some participants (IDs 18 and 49) initially circulated their allocations across the team to get a sense of the members' abilities and then after about 10 iterations converged on the (optimal) allocation. Other participants (ID 41) also circulated their allocations but never converged, whereas some (ID 69) avoided exploration altogether and instead stubbornly stuck to a fixed (suboptimal) choice throughout.

To quantify human performance, we computed allocation probability as a function of iterations for the different experimental conditions (Figure 2A). We found that in all cases the probability of choosing the optimal allocation increased steadily in time and achieved the highest value by the last iteration, even though in some cases (Figure 2A, condition III, right panel) the initial probability was below chance due to participants' general dislike of the auditory task. Indeed, the initial allocation probabilities' 95% CIs for the optimal arms in conditions I, II, and III were [.19, .36], [.16, .32], and [.00, .05], respectively, and in the last iteration they were [.35, .54], [.41, .61], and [.27, .46], with chance level being at .17. Turning next to the accumulated team score, we plotted the resulting curves in Figure 2B along with the UCB and Thompson Sampling predictions (averaged over 1000 runs) and a random baseline. To simulate human agents, we estimated participant success probabilities in each task based on the behavioral data (.70 for lexical, .65 for auditory, and .57 for visual). We see that while initially team performance was at the random allocation level, by the last iteration it significantly exceeded it, and in some cases (condition II; optimal allocation: [lexical, visual, auditory]) it approached the optimal strategies (possibly because that condition was aligned with the participants' native linguistic skills).

4 DISCUSSION

Inspired by Transactive Memory Systems (TMS), we introduced a novel task allocation paradigm whereby participants had to allocate puzzles from different classes between them and their team members to maximize team performance. We found that people deployed different exploration strategies, and were able to successfully integrate information about the strengths and weaknesses of the different team members to infer good task allocations. By further formalizing this problem as a multi-armed bandit, we showed how human performance compared to well-known theoretical strategies, namely, Thomspon Sampling and Upper Confidence Bound.

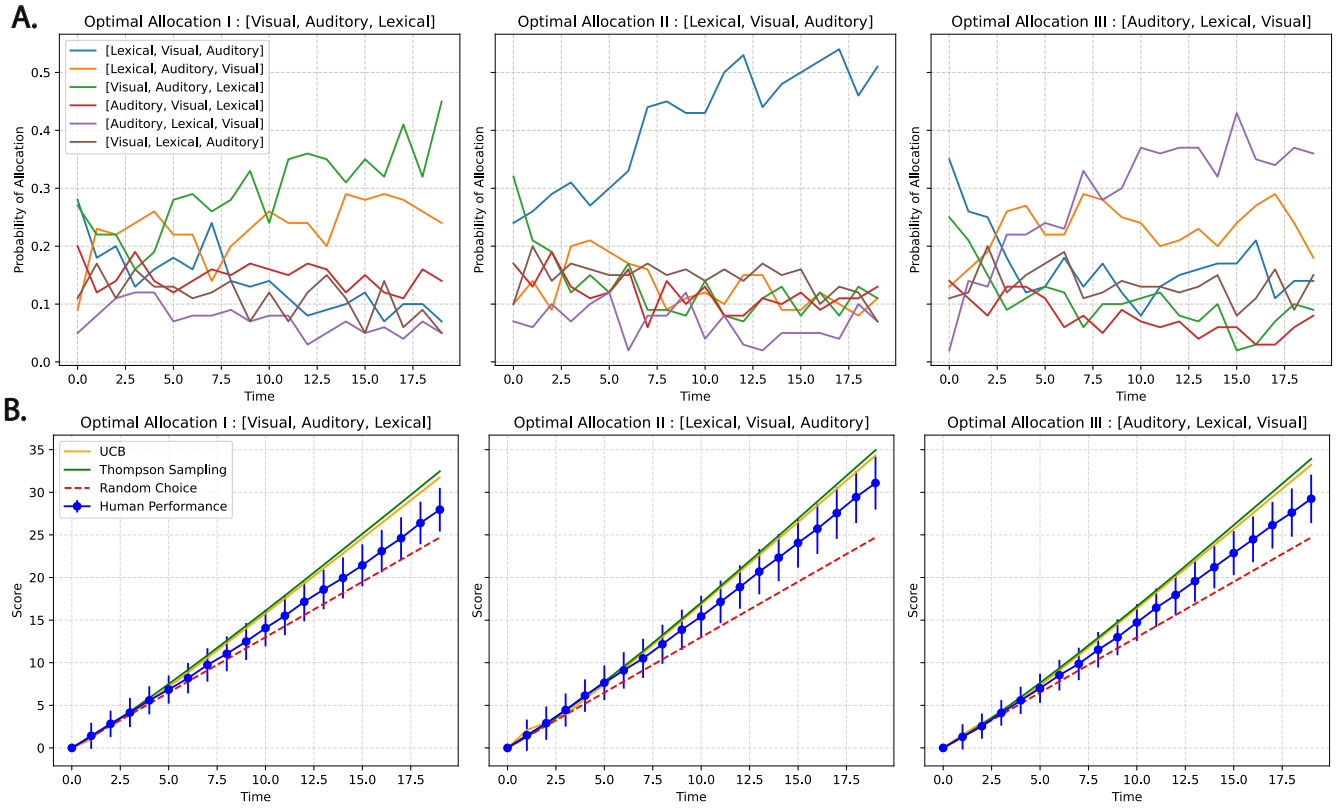

Figure 2: Allocation probabilities and team performance. (A) Choice probabilities as a function of time for the three scenarios considered (an allocation of $[t_0, t_1, t_2]$ means that team member 0 (human) is assigned task t_0, team member 1 (bot) is assigned task t_1, and team member 2 (bot) is assigned task t_2). (B) Average human team performance vs. UCB, Thompson sampling, and random performance as applied to the combinatorial bandit problem. Error bars indicate 95% confidence intervals (CIs).

Our results serve as a first step towards a comprehensive study of coordination in teams. We believe that our paradigm and theoretical framework provide clear avenues toward this goal. First, while it is possible to parametrically explore different bot regimes, we are currently working on including more realistic artificial agents like large language models as well as humans to study human-machine and human-human coordination. Second, our framework can be easily generalized to a multi-agent setup whereby multiple agents allocate tasks among themselves with varying degrees of communication to better align with TMS. Third, an individual-level analysis is likely to be very informative here. By fitting models to individual participants we can characterize the distribution of strategies deployed by participants. We hope to report on all of these directions in the near future.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research (ONR) MURI grant N00014-22-1-2813 to FB and TLG. The authors declare no competing interests.

REFERENCES

[1] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. 2002. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47, 235–256.

[2] Jonathan D Cohen, Samuel M McClure, and Angela J Yu. 2007. Should i stay or should i go? how the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362, 1481, 933–942.

[3] Emilie Kaufmann, Nathaniel Korda, and Rémi Munos. 2012. Thompson sampling: an asymptotically optimal finite-time analysis. In *International conference on algorithmic learning theory*. Springer, 199–213.

[4] Wenjun Mei, Noah E Friedkin, Kyle Lewis, and Francesco Bullo. 2017. Dynamic models of appraisal networks explaining collective learning. *IEEE Transactions on Automatic Control*, 63, 9, 2898–2912.

[5] Joshua C Peterson, Ruairidh M Battleday, Thomas L Griffiths, and Olga Russakovsky. 2019. Human uncertainty makes classification more robust. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9617–9626.

[6] Eric Schulz, Emmanouil Konstantinidis, and Maarten Speekenbrink. 2015. Learning and decisions in contextual multi-armed bandit tasks. In *CogSci*.

[7] Siwei Wang and Wei Chen. 2018. Thompson sampling for combinatorial semi-bandits. In *Proceedings of the 35th International Conference on Machine Learning* (Proceedings of Machine Learning Research). Vol. 80. (Oct. 2018), 5114–5122. https://proceedings.mlr.press/v80/wang18a.html.

[8] Pete Warden. 2018. Speech commands: a dataset for limited-vocabulary speech recognition. *arXiv preprint arXiv:1804.03209*.

[9] Daniel M Wegner. 1987. Transactive memory: a contemporary analysis of the group mind. In *Theories of group behavior*. Springer, 185–208.