Optimizing Decision-Maker's Intrinsic Motivation for Effective Human-AI Decision-Making

Zana Buçinca
zanabucinca@g.harvard.edu
Harvard University
Boston, MA, USA

ABSTRACT

AI advice is increasingly incorporated into decision-making processes, but evidence suggests that decision-makers often struggle to effectively integrate this advice, leading to tendencies to over-rely or under-utilize AI. My research challenges our field's assumption that decision-makers are inherently motivated to engage with AI. I have discovered that cognitive motivation is essential for individuals to actively engage with, critically evaluate, and effectively incorporate AI advice into decision-making. Thus, I propose that AI-powered decision support systems designed to enhance decision-makers' motivation will improve decision-making efficacy. To this end, I have developed two systems that bolster decision-makers intrinsic motivation by supporting their competence and autonomy. Empirical results suggest that fostering intrinsic motivation not only leads to enhanced decision-making performance but also improves the subjective experience when compared to no decision assistance or existing decision support paradigms. This research proposes a paradigm shift in the design of AI-assisted decision-making tools, moving towards systems that improve decision performance via enhancing decision-makers' intrinsic motivation to engage with the task and the decision support.

CCS CONCEPTS

• **Human-centered computing** → **Interaction design**; **Empirical studies in interaction design**; **Systems and tools for interaction design**.

KEYWORDS

decision support, intrinsic motivation, cognitive engagement, human-centered AI, human-AI interaction

ACM Reference Format:
Zana Buçinca. 2024. Optimizing Decision-Maker's Intrinsic Motivation for Effective Human-AI Decision-Making. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3613905.3638179

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3638179

1 INTRODUCTION

AI advice is increasingly being infused into decision-making processes, with the underlying presumption that it will enhance decision-makers' abilities to combine their expertise with AI advice for improved decision outcomes [12]. Yet mounting evidence shows that decision-makers struggle to incorporate AI advice into their decisions, often over- or under-relying on AI [1, 4, 16, 21–24]. Challenging our field's assumption that people will by default engage with AI support, my work has demonstrated that cognitive motivation is a critical factor for individuals to engage with, evaluate, and appropriately incorporate AI advice into decision-making [4]. My work also has shown that the current design and evaluation of AI decision-support tools assumes cognitive engagement and does not support decision-makers' motivation to engage with the provided information [2, 4].

Having demonstrated the importance of cognitive motivation in human-AI decision-making, with my research I now seek to design AI-powered decision support systems that motivate decision-makers to engage with the decision task and with the provided AI support. Specifically, I plan to support decision-makers' *intrinsic* motivation, which stems from internal factors such as personal interest and enjoyment with the task, rather than external forcing or rewards [8]. This form of motivation is associated with achieving high performance and overall well-being across tasks and settings. To do so, I will focus on supporting decision-makers' *competence* (need for ability and confidence) and *autonomy* (need for independence and freedom), as two of the three psychological needs that underlie intrinsic motivation, as posited by the Self-Determination Theory (SDT) [8]. In two separate projects spanning distinct domains, I have already demonstrated how supporting decision-makers' intrinsic motivation and developing AI-driven systems to accommodate decision makers competence and autonomy led to novel systems [5] and interaction techniques [6]. Most importantly, these efforts yielded superior human-AI decision outcomes and decision-making processes compared to existing decision-support paradigms.

My dissertation work conceptually and technically contributes to the AI-assisted decision-making space. I hypothesize that AI decision-support tools, designed to

enhance decision-makers' intrinsic motivation through interaction techniques and explanations, will yield superior objective outcomes and enhance the subjective experience of decision-makers, as compared to existing AI decision-support paradigms. Going forward, I plan to pursue three research directions for building AI that supports human intrinsic motivation.

- *How to intervene?*— I will design novel explanations that support decision-makers' competence, such as by enabling *learning* about the domain or the *discovery* of real-world phenomena related to the decision-making task.
- *When to intervene?* — I will design novel interaction paradigms and build novel computational approaches that select appropriate AI support to optimize different objectives of decision-makers based on the context while supporting their autonomy.
- *What to optimize?* – I will devise technical measures and proxies that capture the higher-level constructs of competence and autonomy in human-AI decision-making.

2 MY WORK SO FAR

2.1 Decision-Maker's Cognitive Motivation Influences Their Engagement With AI in AI-Assisted Decision-Making

Decision-making is a cognitively effortful task and numerous prior studies suggest that people react adversely to the effort required for task accomplishment, be it physical or mental [13, 14]. Unless the task is intrinsically rewarding, individuals typically opt for minimal effort in its execution. I hypothesized that the current paradigm of AI support enabled and exacerbated heuristic thinking (i.e., System 1 thinking) by providing people with readily available decisions they could rely on without spending much effort.

Building upon research in clinical decision-making [17] and upon theories of curiosity [18], I designed a series of *cognitive forcing functions* as interaction design interventions in human-AI decision-making. These *cognitive forcing functions* aimed at overcoming cognitive biases by engaging users in analytical thinking (i.e., System 2 thinking). To illustrate, one of the interventions included asking the person to make the decision on their own prior to seeing the AI recommendation, to provide them with an opportunity to think about their decision and not simply anchor that of the AI. Our results showed that cognitive forcing functions statistically significantly reduced overreliance on AI compared to the standard approach of providing people with AI recommendations and explanations [4]. Yet people disliked these interventions, possibly because these interventions operated via extrinsic motivation. Importantly, I also found that people who were intrinsically motivated to think (i.e., those with higher Need for Cognition [7]) benefited from AI assistance and the interventions more than those with lower cognitive motivation.

Given our field's implicit assumption about human motivation to engage with AI advice, I showed how evaluation tasks and metrics that carry this assumption, yet that have been largely embraced by the research community can, in fact, be misleading in evaluating explainable AI systems [2]. To gauge the effectiveness of explanations, researchers commonly employ proxy tasks, in which participants are asked to predict the AI's decision based on the AI's explanation. I demonstrated that the results of evaluations that use such proxy tasks, where people are explicitly asked to engage with the AI explanations, might not predict the results of actual tasks, in which people make their *own* decisions and can choose whether and how much to attend to the AI. Further, explainable AI systems are often designed with the goal of promoting user trust, which I also showed may not be a predictor of how well people will perform with the system. Because these results have profound implications on the explainable AI landscape, this work was recognized with the Best Paper Award at the Intelligent User Interfaces conference (2020).

2.2 Designing AI-powered Tools to Support Decision-Makers' Intrinsic Motivation

2.2.1 Supporting Competence: Learning Decision-Support Policies to Optimize Decision-Makers' Accuracy and Learning. Competence reflects an individual's desire to feel effective and skilled in their activities [8]. An approach to supporting decision-makers need for competence is for the AI decision-support tools to improve their *learning* about the task and their long-term *skill improvement* along with their immediate decision accuracy. Yet studies show current AI support paradigm of showing recommendations and explanations for every decision, in addition to hindering complementary team performance, impedes learning about the domain [9] and may even contribute to the deskilling of the decision maker in the long term [20]. Evidence suggests that different AI assistance types may benefit learning and accuracy differently for different groups of people. But how to decide which kind of AI assistance is best for whom and when? To address this challenge, in a recent project [6], I cast human-AI decision-making as a reinforcement learning problem, learning optimal policies for selecting AI assistance types that optimize decision-makers accuracy or learning while accounting for relevant contextual factors, including individual differences in motivation to think. Our results show that different learned policies were successful across objectives for different groups of people. Compared to the simple explainable AI approach of showing AI decision recommendations and explanations, people favored

and took pleasure in the task more when using the personalized policies that led to better performance across learning and accuracy. My research demonstrates that supporting and optimizing for decision-makers' intrinsic motivation results in superior objective and subjective measures in AI-assisted decision-making compared to simple explainable AI approaches.

2.2.2 Supporting Autonomy: Designing LLM-Powered Systems to Support Decision-Makers' Autonomy in AI Impact Assessment. Autonomy is the fundamental need for individuals to have control and choice over their actions and decisions [8]. Access to structured information that gives the decision-maker control by enabling informed decision-making is critical for supporting autonomy. In the context of AI development and deployment, AI practitioners frequently encounter intricate decisions laden with values [11], and they must carefully consider how these decisions can impact a wide range of stakeholders. Thus foreseeing the downstream effects of deploying AI systems remains a challenging task for which decision-makers have no support in place. As part of an internship project at Microsoft Research, I developed AHA! (Anticipating Harms of AI), a framework to assist AI practitioners and decision-makers in anticipating the harms of AI systems pre-deployment [5]. Through its structure, AHA! helps organize the problem space so AI practitioners can have the autonomy to make decisions that reflect their values. AHA! systematically considers the interplay between problematic AI behaviors and their potential impacts on stakeholders by narrating these conditions through vignettes. These vignettes are then filled with harm examples by crowds and language models. Examining 4113 harms surfaced by AHA!, we found that AHA! generates meaningful examples of harms, with different scenarios and AI behaviors (e.g., false positives/negatives) resulting in different types of harms. Crowds and language models together generated more diverse harms than either alone. To assess AHA!'s utility, we conducted semi-structured interviews with responsible AI experts. Responsible AI experts discovered meaningful, unexpected harms, valuing AHA!'s systematic approach to surfacing harms and its potential to help them make informed decisions.

3 MY PROPOSED DIRECTION

3.1 *How to intervene?* Designing Novel AI Explanations to Support Human Competence

- **Hypothesis:** The proposed novel explanations will increase decision-maker's competence, resulting in more effective human-AI decision-making compared to existing explanation techniques.
- **Contribution:** Novel AI explanation techniques that support decision-makers' skill improvement.

Current AI explanation design is limited to a few designs (*e.g.,* feature attribution, example-based, saliency maps) and is driven primarily by technical convenience of generating the explanations as opposed to the needs of decision-makers for different tasks and settings [3]. Informed by the competence needs of decision-makers in context, I will design two types of explanations that enable *learning* about the domain and *discovery* of task-related phenomena. While the machine learning community has focused on devising algorithms for contrastive [19] and concept-based explanations [15], whether such explanations are useful to the decision-makers' performance remains unstudied. I will design contrastive and concept-based explanations to enhance people's accuracy and skills in AI-assisted decision-making.

Contrastive explanations. In many contexts, such as for treatment selection in clinical settings, decision-makers have guidelines in place for making decisions. Rather than being presented with explanations that highlight features that contributed to the AI recommendation, in such contexts, decision-makers seek to understand why an AI recommendation differs from the guidelines [10]. I plan to explore the design space of contrastive explanations that highlight the contrast between the AI's recommendation and given guidelines and devise technical approaches to generate useful contrastive explanations. We hypothesize that because contrastive explanations highlight (1) the knowledge gap of the inquirer and (2) are shorter, and thus easier to parse, they will result in improved knowledge acquisition and learning from the decision-maker compared to common explanations that highlight all the factors that contributed to AI's prediction.

Concept-based explanations. In other contexts, decision-makers may have to deal with a large feature space, spanning hundreds or thousands of granular features. Yet there may exist higher-level concepts (*e.g.*, disease) that are correlated with groups of granular features in concert (*e.g.*, lab test results). Existing explanation designs that highlight features or examples are not useful in such instances. Instead of presenting people with granular features, presenting higher-level concepts may be a more useful form of decision support, allowing decision-makers to reason at the level they think about in decision-making. Particularly helpful such explanations would be in domains in which the underlying science is still evolving, such as in antidepressant treatment selection. Clinicians would be able to *learn* from and *discover* underlying mechanisms based on the concepts presented by the AI. I plan to (1) devise novel machine learning approaches that learn such higher-level concepts from lower-level features without labels (which often do not exist), and (2) explore the design space of presenting such concepts to decision-makers.

3.2 *When to intervene*? Designing Novel Interaction Paradigms To Support Autonomy

- **Hypothesis:** The proposed mixed-initiative interaction intervention will lead to increased autonomy, resulting in more effective human-AI decision-making than existing AI decision-support paradigms.
- **Contribution:** Novel mixed-initiative interaction paradigm that supports decision-makers' autonomy.

Learning to adaptively present decision-makers with different AI explanations based on contextual factors is increasingly being seen as a promising solution for AI-assisted decision-making. Yet employing a model-driven approach to present a single type of explanation to decision-makers for a specific decision instance may inadvertently undermine their autonomy since they lack control over the type of support they receive. To enhance decision-makers' autonomy, one effective strategy may be to provide them with the capability to choose among multiple explanation designs presented at the interface level. This feature empowers decision-makers to select AI explanations that best align with their preferences and needs for specific decision instances, giving them greater control over the assistance they receive. I will first investigate the design space of offering multiple explanation design options and how to present those in a non-overwhelming way. Recognizing that selecting assistance for each decision can be tedious, I will also develop personalized recommendation policies. These policies will learn to identify the most beneficial types of support for each decision-maker based on context and their previous selections and prioritize presenting these explanation designs as top options on the interface. This mixed-initiative approach will enable the decision-maker to be in control, while also leveraging the underlying model to guide them towards explanations that may lead to optimal decision outcomes.

3.3 *What to optimize*? Devising Measures and Proxies for Capturing Decision-Makers' *Competence* and *Autonomy* in AI-Assisted Decision-Making

- **Hypothesis:** There exist proxy measures in AI-assisted decision-making that are predictors of decision-makers' *competence* and *autonomy*.
- **Contribution:** Novel metrics that can be used as proxies for optimizing *competence* and *autonomy* in AI-assisted decision-making.

My work has demonstrated that computational models such as RL may be a promising approach to modeling human-AI decision-making and learning policies for selecting appropriate assistance for optimizing different objectives [6]. Yet a challenge in optimizing for intrinsic motivation is that high-level constructs such as *competence* and *autonomy* cannot be easily measured on every decision instance or often enough to obtain reliable signals of improvement. I will examine and identify measures and proxies that are predictors of these constructs in AI-assisted decision-making. For example, time spent on a decision instance may be correlated with cognitive engagement and learning, which in turn are predictors of *competence*. I will conduct large-scale experiments to explore correlations between validated measures of *competence* and *autonomy* from psychology and various proxy indicators that may predict these constructs in the context of AI-assisted decision-making.

4 CONTRIBUTION

In summary, my dissertation has contributed and will further contribute to the conceptual understanding of human-AI decision-making while also introducing innovative metrics, explanations, and interaction techniques designed to optimize the effectiveness of human-AI decision-making *and* the decision-makers' well-being.

ACKNOWLEDGMENTS

The author thanks Krzysztof Gajos, Katy Gero, and Sohini Upadhyay for their valuable insights and feedback on this manuscript. This research was partially supported by the National Science Foundation under Grant No. IIS-2107391 and an IBM PhD Fellowship.

REFERENCES

[1] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 2021. Does the whole exceed its parts? the effect of ai explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[2] Zana Buçinca, Phoebe Lin, Krzysztof Z. Gajos, and Elena L. Glassman. 2020. Proxy Tasks and Subjective Measures Can Be Misleading in Evaluating Explainable AI Systems. In *Proceedings of the 25th International Conference on Intelligent User Interfaces (IUI '20)*. ACM, New York, NY, USA.

[3] Zana Buçinca, Alexandra Chouldechova, Jennifer Wortman Vaughan, and Krzysztof Z. Gajos. 2022. Beyond end predictions: stop putting machine learning first and design human-centered AI for decision support. *NeurIPS Workshop on Human-Centered AI (HCAI)* (2022).

[4] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–21.

[5] Zana Buçinca, Chau Minh Pham, Maurice Jakesch, Marco Tulio Ribeiro, Alexandra Olteanu, and Saleema Amershi. 2023. AHA!: Facilitating AI Impact Assessment by Generating Examples of Harms. *arXiv e-prints* (2023), arXiv–2306.

[6] Zana Buçinca, Siddharth Swaroop, Amanda E. Paluch, Susan A. Murphy, and Krzysztof Z. Gajos. 2023. Offline Reinforcement Learning for Adaptive Support in AI-Assisted Decision-Making. *(under review)* (2023), 1–22.

[7] John T. Cacioppo and Richard E. Petty. 1982. The need for cognition. *Journal of Personality and Social Psychology* 42, 1 (1982), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

[8] Edward L Deci, Anja H Olafsen, and Richard M Ryan. 2017. Self-determination theory in work organizations: The state of a science. *Annual review of organizational psychology and organizational behavior* 4 (2017), 19–43.

[9] Krzysztof Z Gajos and Lena Mamykina. 2022. Do People Engage Cognitively with AI? Impact of AI Assistance on Incidental Learning. In *27th International Conference on Intelligent User Interfaces*. 794–806.

[10] Maia Jacobs, Jeffrey He, Melanie F. Pradier, Barbara Lam, Andrew C. Ahn, Thomas H. McCoy, Roy H. Perlis, Finale Doshi-Velez, and Krzysztof Z. Gajos. 2021. Designing AI for Trust and Collaboration in Time-Constrained Medical Decisions: A Sociotechnical Lens. In *Proceedings of CHI'21*. To appear..

[11] Maurice Jakesch, Zana Buçinca, Saleema Amershi, and Alexandra Olteanu. 2022. How different groups prioritize ethical values for responsible AI. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 310–323.

[12] Ece Kamar, Severin Hacker, and Eric Horvitz. 2012. Combining Human and Machine Intelligence in Large-scale Crowdsourcing. (2012).

[13] Wouter Kool and Matthew Botvinick. 2018. Mental labour. *Nature human behaviour* 2, 12 (2018), 899–908.

[14] Wouter Kool, Joseph T McGuire, Zev B Rosen, and Matthew M Botvinick. 2010. Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General* 139, 4 (2010), 665.

[15] Isaac Lage and Finale Doshi-Velez. 2020. Learning interpretable concept-based models with human feedback. *arXiv preprint arXiv:2012.02898* (2020).

[16] Vivian Lai, Chacha Chen, Q Vera Liao, Alison Smith-Renner, and Chenhao Tan. 2021. Towards a Science of Human-AI Decision Making: A Survey of Empirical Studies. *arXiv preprint arXiv:2112.11471* (2021).

[17] Kathryn Ann Lambe, Gary O'Reilly, Brendan D Kelly, and Sarah Curristan. 2016. Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review. *BMJ quality & safety* 25, 10 (2016), 808–820.

[18] George Loewenstein. 1994. The psychology of curiosity: A review and reinterpretation. *Psychological bulletin* 116, 1 (1994), 75.

[19] Tim Miller. 2018. Contrastive explanation: A structural-model approach. *arXiv preprint arXiv:1811.03163* (2018).

[20] Tapani Rinta-Kahila, Esko Penttinen, Antti Salovaara, and Wael Soliman. 2018. Consequences of Discontinuing Knowledge Work Automation-Surfacing of Deskilling Effects and Methods of Recovery. In *Proceedings of the 51st Hawaii International Conference on System Sciences*.

[21] Max Schemmer, Patrick Hemmer, Niklas Kühl, Carina Benz, and Gerhard Satzger. 2022. Should I follow AI-based advice? Measuring appropriate reliance in human-AI decision-making. *arXiv preprint arXiv:2204.06916* (2022).

[22] Siddharth Swaroop, Zana Buçinca, Krzysztof Z. Gajos, and Finale Doshi-Velez. 2024. Accuracy-Time Tradeoffs in AI-Assisted Decision Making under Time Pressure. In *29th International Conference on Intelligent User Interfaces (IUI '24)*. ACM.

[23] Helena Vasconcelos, Matthew Jörke, Madeleine Grunde-McLaughlin, Tobias Gerstenberg, Michael S Bernstein, and Ranjay Krishna. 2023. Explanations can reduce overreliance on ai systems during decision-making. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–38.

[24] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. 2019. Understanding the effect of accuracy on trust in machine learning models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.