An Experimental Method for Studying Complex Choices

Nikolos Gurney[1] [0000−0003−3479−2037], Tyler King[2] and John H. Miller[3,4]

[1] University of Southern California, Los Angeles CA 90094, USA
gurney@ict.usc.edu
[2] Cornell University, Ithaca NY 14850, USA
ttk22@cornell.edu
[3] Carnegie Mellon University, Pittsburgh PA 15213, USA
JM7T@andrew.cmu.edu
[4] Santa Fe Institute, Santa Fe NM 87501, USA

Abstract. The promise of computational decision aids, from review sites to emerging augmented cognition technology, is the potential for better choice outcomes. This promise is grounded in the notion that we understand human decision processes well enough to design useful interventions. Although researchers have made considerable advances in the understanding of human judgment and decision making, these efforts are mostly based on the analysis of simple, often linear choices. Cumulative Prospect Theory (CPT), a famous explanation for decision making under uncertainty, was developed and validated using binary choice experiments in which options varied on a single dimension. Behavioral science has largely followed this simplified methodology. Here, we introduce an experimental paradigm specifically for studying humans making complex choices that incorporate multiple variables with nonlinear interactions. The task involves tuning dials, each of which controls a different dimension of a nonlinear problem. Initial results show that in such an environment participants demonstrate classic cognitive artifacts, such as anchoring and adjusting, along with falling into exploitive traps that prevent adequate exploration of these complex decisions. Preventing such errors suggest a potentially valuable role for deploying algorithmic decision aids to enhance decision making in complex choices.

Keywords: Complex choice · Judgment and decision making · Behavioral economics · Decision aids.

1 Introduction

From buying groceries to the purchase of a house, and from choosing a course schedule to accepting a job offer, individuals often confront complex choices. The necessity to integrate information across multiple dimensions is a key feature of these choices. For example, in the purchase of a car, one must pick how the car is powered, how many passengers it can carry, expected mileage, top speed, quality, price, and so on. If the dimensions do not interact, then choices are relatively easy as they can be made one dimension at a time. Such linear scenarios, however, are typically the exception rather than the rule. Most choice dimensions have nonlinear interactions (for example, to carry more passengers you need a larger car and that impacts the performance of the motor etc.) We investigate this realm of decision making over complex choices to both explore how individuals approach them as well as suggest a productive role for computational assistance.

We consider a choice complex when it involves multiple variables with nonlinear interactions. This includes mundane choices, such as deciding what groceries to buy, as well as monumental, i.e. high resource, choices over cars, houses, and life events from jobs to marriage. Given the nature of complex choices, there may be a role for computational decision aids (CDAs), i.e. digitally mediated information sources that are frequently, but not always, supported by machine learning. CDAs can improve choice making in these complex environments by implementing insights from behavioral science (e.g. [2,9,10]). A growing body of research also examines how people make use of CDAs once they are deployed, how they impact choices, and even how to regulate them (e.g. [1,3,7]). However, the underlying theories of human judgment and decision making that inform the design of CDAs are almost entirely grounded in the analysis of simple, often linear, choices. For example, Cumulative Prospect Theory (CPT) was developed and validated using binary choice experiments in which options varied on a single, linear dimension [13]. CPT and the behavioral predictions that emerge from it are widely cited, applied, and adapted in CDA research (e.g. [4, 6, 11]) even though choices rarely look like those from the original research.

2 Methodology

Our methodology generates continuous, n-dimensional search spaces that are explored using a digital interface. We take inspiration from complexity science and adapt the *fitness landscape* metaphor to represent choice sets. Fitness landscapes were originally used as a means of describing the relationship between the genetics of an organism, represented by a location in its "fitness landscape," and its success in the world, represented by the elevation of that location [14]. If the landscape is relatively smooth, then it is easy for slight, uni-dimensional changes to result in climbs to higher elevation. As the landscape becomes more rugged, that is, more nonlinear, the impact on elevation of moving along a single dimension becomes much harder to predict. Such landscapes are characterized by many local optima, with numerous peaks and valleys, that increase the difficulty of finding the highest peak. Classically, if one wants to climb to the top of a mountain in a fog, it is far easier to climb Mount Fuji than a rugged expanse of, say, the Rocky Mountains.

A popular generalized formal model of fitness landscapes, the NK model [5], has been used to describe complex settings from biology [12] to business operations [8]. The basic concept of the NK model is that each of an object's N dimensions contributes to its overall fitness, but that contribution is contingent on K interactions with the other dimensions.

2.1 Procedural Generation of Landscapes

We use a proof-of-concept model for procedurally generating N-dimensional choice landscapes.[5] The landscapes that we generate are continuous over their edges, meaning that if, say, you go above the upper limit of a dimension you are placed at its lower limit. A two-dimensional landscape would thus wrap its upper side to its lower, and its left side to its right—equivalent to an ant wandering around on a floating donut (aka a torus).

We generate a set of elevation points for such a landscape. Here, we insure a certain degree of continuity between adjacent points, though this is easily changed. First, we define the underlying parameters for the landscape, for example, its width, height, maximum elevation, number of "peaks" (local maxima), and the steepest acceptable slope between two points. Next, we randomly place the peaks in the landscape and check that they are in compatible locations, namely, that they are far enough apart that one peak will not be subsumed by another. Finally, a smoothing algorithm fills in the remaining points in the landscape. This algorithm ensures that all of the peaks are separated by intervening "valleys."

2.2 User Interface

Users move about a landscape by setting rotating dials akin to those found on combination locks or analog radios (see figure 1). Once the user has decided on the dial settings, the system is queried and the user receives a report of the landscape's elevation at that setting. Thus, two dials are needed to explore a 3-dimensional landscape, one of which can be thought of as controlling the east-west dimension (longitude) while the other determines the north-south dimension (latitude). While we implemented discrete dials, continuous-valued dials are also possible. The reported elevation, captured in an onscreen history of past dial settings and their associated elevation, can be manipulated in various ways, for example, it can have added noise, be re-scaled so that elevations are negative (to explore loss aversion), and so on.

3 Experimental Design

Participants are tasked with tuning the available dials to achieve a stated (and incentivized) goal. The goals are designed for gaining insights about decision making in such environments. For example, maximizing (or minimizing in a loss domain) the final elevation. Such goals are motivated by an associated narrative and incentive structure. For example, when trying to recover economic choice behavior, researchers can introduce an incentive structure in which participants earn more rewards for discovering better dial settings.

[5] A tutorial is available at http://nmgurney.com/complex-choice-landscape-maker/

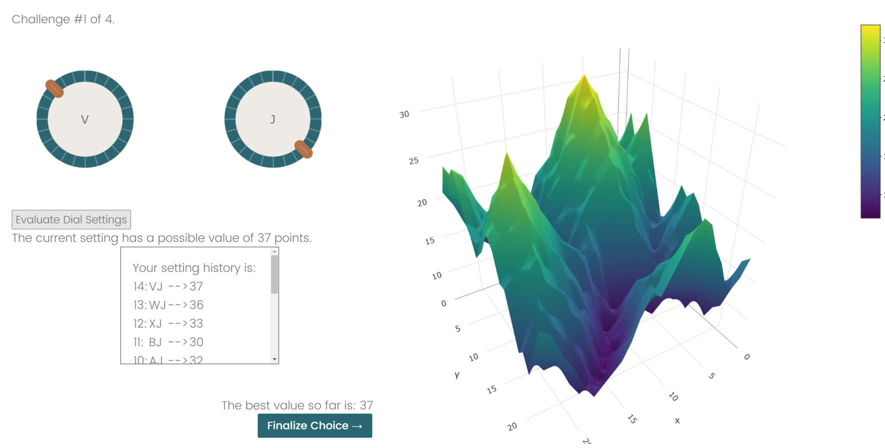

Fig. 1. The user interface is presented on the left. Note the two dials with orange handles. Users click and drag the handles to adjust the dials. The "Evaluate Dial Settings" button allows users to get feedback on a dial setting. The dial setting history is displayed below as well as a submission button, which finalizes their effort on a task. The right image is a rendering of an underlying 3-dimensional landscape. The global optimum is the yellow peak set in the middle-top and the minimum is located in the dark blue valley in the middle bottom of the image.

3.1 Pilot Study Design

302 participants (Mechanical Turk workers) each completed nine dial tuning tasks for a total of 2718 observations. Each of the nine tasks had a unique three-dimensional, discrete search space comprised of 24 points in each tunable dimension, implying 576 possible locations. The search spaces (landscapes) varied in their ruggedness, having either one, two, or four peaks. Metaphorically, a landscape with four peaks is more rugged than one with two peaks, and thus finding, say, the global maximum should be more challenging. The experiment design crossed each level of ruggedness with three levels of feedback accuracy: no noise, low noise, and high noise. When a task provided noisy feedback, a random integer on a fixed interval (either [-1,1] or [-3,3]) was drawn and added to the underlying elevation value each time the dial settings were checked. For each task, a participant could tune the dials and sample unique locations as many times as they wished. The last dial tune submitted became their chosen setting for a given task. The order of the nine different tasks for each participant was randomized. Each participant was paid a bonus based on her or his performance in one of the nine tasks (drawn at random) in addition to a base payment for participation. To further incentivize effort, each participant's performance across all nine tasks was totaled and became entries in a lottery for one of five large cash prizes. Finally, note that participants were not informed of the global maximum that they could obtain in a given task, so such information would have to be gained during the course of exploration.

3.2 Results

Please note that we did not formalize any hypotheses prior to data collection, thus we view these results as exploratory and present them as a means of demonstrating the ability of the methodology to generate behavioral insights. Additionally, we report linear models for simplicity and leave more robust approaches to future work.

A multiple linear regression model revealed evidence supporting the notion that the final value that participants submitted for the initial dial tuning task served as an aspiration, or anchor, level in the subsequent tasks ($F(9, 2408) = 70.40, p < 0.001, R2 = 0.21, \beta_{anchor} = 0.394, p < 0.001$). We also analyzed the effect of explore and exploit strategies on attainment in the first round and subsequent rounds. Performance in the first round was highly correlated with exploration ($F(9, 292) = 10.54, p < 0.001, R2 = 0.22, \beta_{explore} = 0.670, p < 0.001$), which we define as a location shift of greater than two positions (Manhattan distance) as was performance across missions ($F(9, 2408) = 41.10, p < 0.001, R2 = 0.22, \beta_{explore} = 0.531, p < 0.001$). In general, participants tended to fall into exploitive traps that, by limiting exploration early on in the process, limited their potential gains. However, when participants used a strategy of exploring early and then later exploiting local information in a given task, they tended to perform much better ($F(9, 2408) = 39.32, p < 0.001, R2 = 0.12, \beta_{exploreExploit} = 4.485, p < 0.001$).

Across our models we controlled for landscape complexity and consistently saw that the 2-peaked landscape was correlated with worse performance than both the 1-peaked and 4-peaked landscapes. Further investigation revealed that this is likely an artifact of the fixed landscape size, as adding more peaks increases the average value in the plain. This means that, without correcting for average elevation, performance on a 4-peaked landscape looks better because participants are simply more likely to land on a higher location than the other landscapes. An ANOVA ($F(2, 293) = 4.192, p = 0.016$) looking at performance in the first task with performance adjusted for average elevation, for example, suggests that the 4-peaked landscapes were, in fact, the hardest. A Tukey's HSD test for multiple comparisons found that the mean value of corrected submissions was significantly lower for 4-peaked than 1-peaked landscapes ($p = 0.014, 95\% C.I. = [-5.155, -0.463]$).

4 Conclusion

Most of the choices we face in the world involve multiple, interacting dimensions. Such choices are fundamentally difficult because movement across one dimension can alter how prior decisions across other dimensions impact the value of the choice. We found that when facing such choices, participants fall into previously identified decision failures such as anchoring. Moreover, many of the subjects tended to focus far more on the exploitation of what was known rather than on the exploration of the unknown—which is particularly damaging early on when little is known about the choice landscape. While such a strategy works well in a linear world of single-peaked landscapes, it fails in nonlinear worlds with multiple peaks.

The paradigm we develop provides a tractable way to begin to explore human decision making in the realm of complex choice. We argue that complex choices tend to be the norm for the real-world problems that face most individuals. Understanding better how humans confront such choices is not only important, but essential to developing a more complete appreciation for the human condition. More pragmatically, the work also suggests a potential role for deploying computational decision aids to improve outcomes. For example, such aids can help users avoid the usual decision traps caused by faulty heuristics or search strategies. People have, for millennia, successfully navigated complex choices. Studying what allows us to succeed in the face of such choices will empower us to build resources to continue to flourish in an increasingly complex world.

References

1. Cai, C.J., Winter, S., Steiner, D., Wilcox, L., Terry, M.: " hello ai": uncovering the onboarding needs of medical practitioners for human-ai collaborative decision-making. Proceedings of the ACM on Human-computer Interaction **3**(CSCW), 1–24 (2019)
2. Dietvorst, B.J., Simmons, J.P., Massey, C.: Algorithm aversion: people erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General **144**(1), 114 (2015)
3. Green, B., Chen, Y.: The principles and limits of algorithm-in-the-loop decision making. Proceedings of the ACM on Human-Computer Interaction **3**(CSCW), 1–24 (2019)
4. Grgic-Hlaca, N., Engel, C., Gummadi, K.P.: Human decision making with machine advice: An experiment on bailing and jailing. Proceedings of the ACM on Human-Computer Interaction **3** (2019)
5. Kauffman, S.A., Weinberger, E.D.: The nk model of rugged fitness landscapes and its application to maturation of the immune response. Journal of theoretical biology **141**(2), 211–245 (1989)
6. Lee, M.C., Park, J.: There is no perfect evaluator: An investigation based on prospect theory. Human Factors and Ergonomics in Manufacturing & Service Industries **28**(6), 383–392 (2018)
7. Lee, M.K., Jain, A., Cha, H.J., Ojha, S., Kusbit, D.: Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. Proceedings of the ACM on Human-Computer Interaction **3**(CSCW), 1–26 (2019)
8. Levinthal, D.A.: Adaptation on rugged landscapes. Management science **43**(7), 934–950 (1997)
9. Logg, J.M., Minson, J.A., Moore, D.A.: Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes **151**, 90–103 (2019)
10. Poursabzi-Sangdeh, F., Goldstein, D.G., Hofman, J.M., Wortman Vaughan, J.W., Wallach, H.: Manipulating and measuring model interpretability. In: Proceedings of the 2021 CHI conference on human factors in computing systems. pp. 1–52 (2021)

11. Quinn, P., Cockburn, A.: Loss aversion and preferences in interaction. Human–Computer Interaction **35**(2), 143–190 (2020)
12. Romero, P.A., Arnold, F.H.: Exploring protein fitness landscapes by directed evolution. Nature reviews Molecular cell biology **10**(12), 866–876 (2009)
13. Tversky, A., Kahneman, D.: Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and uncertainty **5**(4), 297–323 (1992)
14. Wright, S., et al.: The roles of mutation, inbreeding, crossbreeding, and selection in evolution (1932)