Learning progress and uncompensated rewards as motivational drivers of engagement

Franziska Brändle (franziska.braendle@tuebingen.mpg.de)
MPI for Biological Cybernetics, Max-Planck-Ring 8
Tübingen, 72076 Germany

Charley M. Wu (charley.wu@uni-tuebingen.de)
Human and Machine Cognition Lab, University of Tübingen, Maria-von-Linden-Str. 6
Tübingen, 72076 Germany

Eric Schulz (eric.schulz@tuebingen.mpg.de)
MPI for Biological Cybernetics, Max-Planck-Ring 8
Tübingen, 72076 Germany

Abstract

Theories of motivation describe how behavior is driven by different factors like external rewards or inherent satisfaction. In this project, we looked at two components of motivation driving people to play games: "fun" — defined as improving one's model of the environment — and the magnitude of available points (without monetary compensation). Here, we test this theory by predicting that engagement is influenced by two factors: fun, which is maximal when learning progress is maximal — corresponding to an intermediate level of difficulty — and the magnitude of point values. We test our predictions in a grid exploration task, in which we manipulate the underlying spatial distribution as well as the magnitude of outcomes. Both participants' behavior and model-based analyses confirmed our predictions.

Keywords: curiosity; fun; intrinsic motivation; exploration; learning progress; reward

Introduction

Research on motivation has looked at a plethora of different concepts (Cogliati Dezza, Schulz, & Wu, 2022) — from extrinsic rewards like monetary compensation (Braver et al., 2014; Murayama, Matsumoto, Izuma, & Matsumoto, 2010) to intrinsic motivation like uncertainty reduction (Berlyne, 1962) and empowerment (Brändle, Stocks, Tenenbaum, Gershman, & Schulz, 2022). Here, we study two forms of motivation that seem to be key drivers of why people play games: fun and scoring points.

"Fun" is a broad concept with many interpretations. Here, we use a computational framework describing "fun" as a measure of progress in learning one's model of the world (Schmidhuber, 2010)[1]. Thus, we can think of fun as accomplishing the rational goal of improving learning progress, which we expect to produce the highest engagement for intermediate difficulties (Vygotksy, 1978; Kidd, Piantadosi, & Aslin, 2012; Wilson, Shenhav, Straccia, & Cohen, 2019; Geana, Wilson, Daw, & Cohen, 2016; Ten, Kaushik, Oudeyer, & Gottlieb, 2021).

People are also motivated to play games by scoring points or achieving high scores (Johnson et al., 2018), even when it carries no monetary payoffs. People are even willing to spend money at arcades or buy virtual game items to get the highest score. One reason might be that uncompensated points share similar reward learning properties with classic extrinsic rewards (Murayama, 2022).

In the current work, we examine both fun and uncompensated points as motivational factors of engagement, in a task where we can control the difficulty and the magnitude of point values. We hypothesized that both factors will influence engagement, producing the highest levels when difficulty is intermediate and when the uncompensated point values are highest.

Experiment

We adopted the spatially correlated multi-armed bandit paradigm (Wu, Schulz, Speekenbrink, Nelson, & Meder, 2018), but greatly increased the size of the environment to 30 x 30 grids (Fig. 1). Participants iteratively sampled tiles on a grid, and observed their point value along with a corresponding color (darker for more points). At any point in the task after the first five trials, participants could decide that they were finished exploring the current grid, and progress onto a new grid. The experiment deterministically ended after 10 minutes, independent of the number of grids or cells participants explored. Participants were not given any performance-dependent bonus to focus on intrinsic motivation.

To manipulate *difficulty*, we changed the smoothness (i.e., spatial correlation) of point value by changing the length-scale parameter $\lambda \in [2^{-2}, \ldots, 2^{4}]$ of a radial basis function kernel used to generate the environments. When λ is small (e.g. 0.25), we generate "rough" environments (Fig. 1a), where the value of a tile provided relatively little information about the values of adjacent tiles. When λ is large, we generate "smooth" environments (Fig. 1b), where values of tiles were strongly predictive of nearby tiles. We manipulated *uncompensated*

[1] Other approaches (e.g. Oudeyer, Gottlieb, & Lopes, 2016) are similarly defined and make comparable predictions.

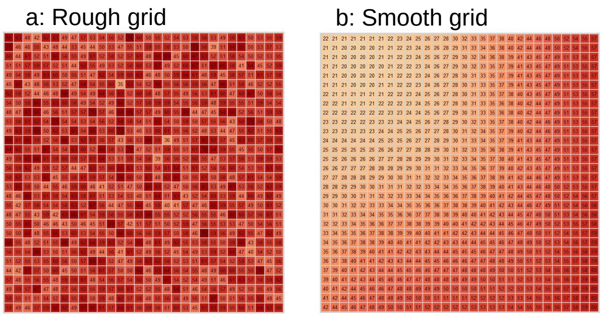

Figure 1: Two example grids with different length-scale λ. **a**: A rough grid with λ = 0.25. **b:** A smooth grid with λ = 16.

points (i.e., point values of the tiles) via the range of point values on a grid, where we uniformly sampled a minimum value between 5 and 35 and set the maximum to the minimum + 40.

We estimated motivation as a function of engagement, based on how many tiles participants explored on each grid. We hypothesized that engagement is influenced by two factors. First, we predicted an inverse-U-shaped relationship between smoothness and engagement, with intermediately smooth grids yielding the highest level of engagement. For both very smooth and very rough grids, there is limited opportunity for learning progress. This is because participants will either very rapidly learn the structure after only a few samples with minimal benefits for further exploration (very smooth) or fail to learn the structure altogether due to a lack of any point value correlation (very rough). In contrast, intermediately smooth grids were expected to induce the greatest levels of engagement because participants could make meaningful learning progress over an extended period. Second, we predicted that engagement and the magnitude of underlying point values would have a linear relationship: a higher magnitude of revealed points should induce greater engagement.

Simulations

We first conducted a series of simulations to test the learning progress hypothesis in our paradigm with a Bayesian learning model. We did not conduct any simulations on the influence of the magnitude of point values, as they would by definition display a simple linear relationship between magnitude and the number of samples. To investigate the influence of the smoothness of the grids more formally, we simulated 2000 grids per value of the length-scale parameter λ. In each grid, we sampled randomly one tile after another. We created a Bayesian model that — with each new sample — updated its predictions over the values of all tiles, based on what was sampled so far. We compared the model predictions to the original values of the grid and calculated the mean squared error. We then took the mean over all grids of each λ separately for each time step to plot the learning progress — a reduction of the error — over time for each kind of environment. As can be seen in Figure 2a, in environments with small λ — rough grids — the model learns fast, but plateaus quickly at an error of around 0.025. In environments with high λ — smooth grids — the model also learns fast and approaches 0 very quickly. Only in environments with intermediate λ values does the model learn over a longer period, which supported our hypothesis. To simulate how long the models would interact with a grid before going on to the next one, we set a threshold for learning progress — the lower limit of how much progress the model would count as "fun". Here, we set the value to 0.001 as this puts the predicted number of interactions in the same range as participants' actual interactions. We counted the number of tiles each model sampled before its absolute learning progress was below this threshold. As expected, we found an inverted U relationship between the λ values and the number of samples (Fig. 2b).

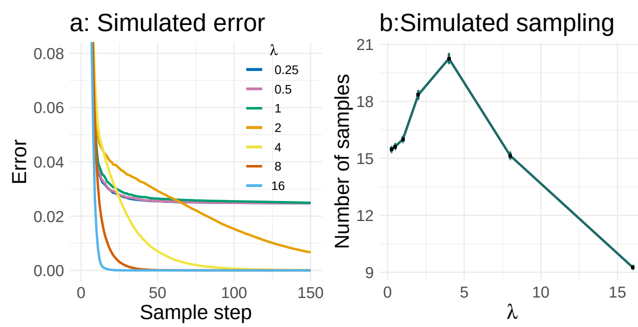

Figure 2: Simulation results. a: Errors of the different models per sample step. Models with small λ quickly plateau and models with high λ quickly approach 0. Models with intermediate λ learn over a longer period. b: Predicted number of interactions per model depending on the error. The model with $\lambda = 4$ learns for the most number of samples. Error bars indicate the standard error of the mean.

Results

We analyzed data from 44 participants recruited via Amazon Mechanical Turk. Participants interacted with an average of 40 grids ($SD = 37.16$) and explored an average of 19 tiles per grid ($SD = 29.49$). We compared how long they engaged with each grid as a function of the grid's smoothness and the magnitude of the underlying point values (Fig. 3). We used a negative binomial mixed-effects regression analysis including a random intercept and found that the λ parameter had a significant positive linear effect ($\beta = 3.44$, $z = 3.39$, $p < .001$) as well as a significant negative quadratic effect ($\beta = -1.57$, $z = -2.27$, $p = .023$), which accounted for the inverted U-shape, while the magnitude had a significant positive effect ($\beta = 2.18$, $z = 4.03$, $p < .001$), confirming our hypotheses.

Discussion

We tested the influence of learning progress and uncompensated points on engagement in a grid search paradigm. We showed that engagement (i.e., the number of samples of each grid) had a negative quadratic relationship with the smoothness of the grids and a linear relationship with the magnitude

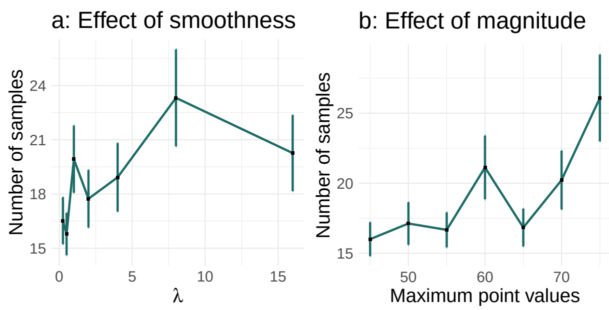

Figure 3: Results a: Average number of samples dependent on the smoothness — the higher λ the smoother the grid. b: Average number of samples dependent on the magnitude of point values (with magnitude binned in bins of size 5). Error bars indicate the standard error of the mean.

of point values: participants liked to interact with intermediately smooth grids — leading to higher learning progress — and high point values the most.

While we were able to develop a model that showed similar behavior as participants, we assumed that they use a random sampling strategy. However, we know that humans of all ages do not sample randomly in this paradigm (Giron et al., 2022). In future work, we would like to implement different sampling strategies to see how they change our model's predictions.

We also assumed that people have the same prior expectations as the actual smoothness of the grids. However, we found in some additional simulations that the favorite "difficulty" level — the peak of the inverse U-shaped curve — can be influenced by participants' priors. We plan on investigating this more closely by studying how the manipulation of priors — showing participants opened grids with different smoothness before the experiment — might lead to different preferences.

Acknowledgments

This work was supported by the Max Planck Society, the Volkswagen Foundation, and a Jacobs Research Fellowship. CMW and ES are supported by the German Federal Ministry of Education and Research (BMBF): Tübingen AI Center, FKZ: 01IS18039A, and funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC2064/1–390727645.

References

Berlyne, D. E. (1962). Uncertainty and epistemic curiosity. *British Journal of Psychology*, *53*(1), 27–34.

Braver, T. S., Krug, M. K., Chiew, K. S., Kool, W., Westbrook, J. A., Clement, N. J., ... others (2014). Mechanisms of motivation–cognition interaction: challenges and opportunities. *Cognitive, Affective, & Behavioral Neuroscience*, *14*, 443–472.

Brändle, F., Stocks, L. J., Tenenbaum, J., Gershman, S., & Schulz, E. (2022). Intrinsically motivated exploration as empowerment. *psyArXiv Preprint*.

Cogliati Dezza, I., Schulz, E., & Wu, C. M. (Eds.). (2022). *The drive for knowledge: The science of human information-seeking*. Cambridge University Press. doi: https://doi.org/10.1017/9781009026949

Geana, A., Wilson, R. C., Daw, N. D., & Cohen, J. D. (2016). Boredom, information-seeking and exploration. *Cognitive Science*.

Giron, A. P., Ciranka, S. K., Schulz, E., van den Bos, W., Ruggeri, A., Meder, B., & Wu, C. M. (2022, Apr). *Developmental changes resemble stochastic optimization.* PsyArXiv. Retrieved from psyarxiv.com/9f4k3 doi: 10.31234/osf.io/9f4k3

Johnson, D., Klarkowski, M., Vella, K., Phillips, C., McEwan, M., & Watling, C. N. (2018). Greater rewards in videogames lead to more presence, enjoyment and effort. *Computers in Human Behavior*, *87*, 66-74.

Kidd, C., Piantadosi, S., & Aslin, R. N. (2012). The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLOS ONE*.

Murayama, K. (2022). A reward-learning framework of knowledge acquisition: An integrated account of curiosity, interest, and intrinsic–extrinsic rewards. *Psychological Review*, *129*, 175-198.

Murayama, K., Matsumoto, M., Izuma, K., & Matsumoto, K. (2010). Neural basis of the undermining effect of monetary reward on intrinsic motivation. *Proceedings of the National Academy of Sciences*, *107*(49), 20911–20916.

Oudeyer, P.-Y., Gottlieb, J., & Lopes, M. (2016). Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. *Progress in Brain Research*.

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, *2*(3), 230-247.

Ten, A., Kaushik, P., Oudeyer, P.-Y., & Gottlieb, J. (2021, Oct 13). Humans monitor learning progress in curiosity-driven exploration. *Nature Communications*, *12*(1).

Vygotksy, L. S. (1978). Mind in society: the development of higher psychological processes.

Wilson, R. C., Shenhav, A., Straccia, M. A., & Cohen, J. D. (2019). The eighty five percent rule for optimal learning. *Nature Communications*.

Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., & Meder, B. (2018). Generalization guides human exploration in vast decision spaces. *Nature Human Behaviour*, *2*, 915—924. doi: 10.1038/s41562-018-0467-4