Intrinsically Motivated Exploration as Empowerment

Franziska Brändle[1,+,*]**, Lena J. Stocks**[1,+]**, Joshua B. Tenenbaum**[2]**, Samuel J. Gershman**[3]**, and Eric Schulz**[1]

[1]Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany
[2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139
[3]Department of Psychology, Harvard University, Cambridge, MA 02138
[*]franziska.braendle@tuebingen.mpg.de. Max Planck Ring 8. 72076 Tübingen.
[+]these authors contributed equally to this work

ABSTRACT

Studies of human exploration frequently cast people as serendipitously stumbling upon good options. Yet these studies may not capture the richness of exploration strategies that people exhibit in more complex environments. We study human behavior in a large data set of 29,493 players of the richly-structured online game "Little Alchemy 2". In this game, players start with four elements, which they can combine to create up to 720 complex objects. We find that players are driven to create objects that empower them to create even more objects. We find that this drive for empowerment is eliminated when people play a version of the game that lacks recognizable semantics, indicating that they use their knowledge about the world to guide their exploration. Our results suggest that the drive for empowerment may be a potent source of intrinsic motivation in richly structured domains, particularly those that lack explicit reward signals.

Introduction

Exploration—seeking out potentially useful information—is prevalent in our everyday lives. From choosing a restaurant to finding a suitable workplace, we need to explore our options to be able to make good decisions. A fundamental tension in all these scenarios exists between exploring unknown options and exploiting known options. An algorithmic account of human exploration must explain both *what* to explore and *when* to explore.

Psychologists and neuroscientists have extensively studied human exploration in simple and highly controlled multi-armed bandit tasks[1,2]. In these tasks, participants choose between a set of options ("arms"), each associated with an unknown reward distribution. It is the participants' goal to maximize rewards by repeatedly sampling arms and collecting the resulting rewards. Ideal agents should explore by combining the immediate reward and the value of information for each action; they can do so by thinking through all possible future actions and calculating how much rewards would increase if more knowledge about the reward distributions was collected. However, such optimal exploration strategies are computationally intractable. Researchers have therefore focused on the heuristic strategies of exploration that humans might employ[3,4]. Some evidence suggests that people use sophisticated uncertainty-based heuristics[5–7].

We propose that human exploration strategies are richer than what has previously been described. In particular, we believe that current models of human exploration do not capture the intrinsically motivated exploration strategies observed in the real world[8–10]. As an example, consider how children play with their environment, curiously trying out new things in order to understand and learn about the world; or how scientists explore and arbitrate between different hypotheses to advance our collective knowledge. In many of these settings, direct rewards are very sparse and it is often not even clear what the reward is. Yet people can spend time on activities without such rewards; these preferences reflect intrinsic exploratory drives. Current laboratory tasks are not rich enough to study these types of behaviors quantitatively. We, therefore, propose to study human exploration in more complex and richly-structured environments. One such environment is the online game *"Little Alchemy 2"*, in which players start out with four basic elements: water, fire, earth, and air. They can then use their intuitive semantic understanding and always combine two elements, which sometimes leads to new elements. Each created element is added to an inventory for use in future combinations (see Fig. 1a). The combination results are semantically meaningful (e.g., combining water with fire produces steam, and can lead to increasingly complex elements, such as humans; see Fig. 1b). Game play is not random: people selectively choose which elements to combine, and thereby follow particular paths through the vast state space of element inventories. Importantly, players do not receive any extrinsic rewards during the game, yet may play for several hours. Thus, we believe that "Little Alchemy 2" offers a better and more realistic testbed to investigate intrinsic exploration strategies than many current laboratory tasks. In the current paper, we analyzed a large data set of 29,493 players who collectively produced more than 4 million trials.

We show that players' exploration behavior is best described by an exploration as empowerment model that we propose in this paper, with an additional contribution of uncertainty-guided exploration. Uncertainty-guided exploration is a well-known strategy that can be formalized as the tendency to combine elements that have not frequently been used before. Exploration as empowerment is a novel description of human exploration that can be formalized as the attempt to create elements that can be used to create even more elements. This is similar to how scientists explore when they are trying to gain insights that enable them to gain even further insights and therefore explore even more. Using two simpler versions of the "Little Alchemy 2" testbed, we show that our previous results can be replicated in an experimental setting and that the effect of empowerment on participants' exploration strategies vanishes if we remove the semantics of the game. These results push our understanding of human exploration strategies away from simple strategies of exploration in simple tasks and towards the rich repertoire of intrinsic exploration strategies found in rich environments.

Extending Models of Human Exploration

Previous studies on human exploration have coalesced around two strategies: random and directed exploration. Both use uncertainty about the available options to guide exploration behavior but differ in how uncertainty is assumed to guide behavior[1]. Whereas directed exploration applies an information bonus to seek out options with higher uncertainty, random exploration predicts that choice stochasticity increases with higher uncertainty across all available options. While earlier studies did not produce consistent empirical evidence for uncertainty-guided exploration in human decision making (e.g.[3, 11]), recent studies have provided converging evidence in favor of such strategies[4, 6, 12, 13].

What many of the previous studies on human exploration have in common, is that they used the fairly simple paradigm of multi-armed bandits and only collected data from a small number of participants. Although these tasks have contributed to a deeper understanding of human exploration behavior, their simplicity might have masked more sophisticated strategies that people could apply in richer settings. Indeed, the strategies humans *can* employ in exploration tasks –and which can be found empirically– are clearly limited by the complexity of the used experimental paradigms[8]. The study of empowerment, for example, requires a change of influence on future options, which is not possible to assess in multi-armed bandits without changing rewards or dynamic states, as well as an intuitive understanding of which actions can be empowering, for example by using an intuitive understanding of which objects in a game can be combined.

To set the stage for our analyses of people's playing behavior, we first describe the "Little Alchemy 2" game in more detail before then explaining the algorithmic ideas behind uncertainty-guided exploration and exploration as empowerment.

A Quintessential Game of Exploration

In the present work, we look at the game "Little Alchemy 2", created and released by Jakub Koziol in 2017. By August 2021, the game has been downloaded over ten million times[14]. The idea of the game is simple: players start with an inventory of only four elements: earth, fire, water, air. Players can create new elements by always combining two already existing elements. The resulting elements are added permanently to the inventory and can be used from then onward (see Fig. 1a). The successful combinations and their results are semantically meaningful. For example, the combination of fire and earth leads to lava, which can be combined with sea to create primordial soup. These can be the first steps to create life and –eventually– human in the further course of the game (see Fig. 1b). "Little Alchemy 2" offers a total of 720 elements, ranging from basic items like energy or glass to extremely specific elements like cookie dough or Frankenstein's Monster. Between these elements, there are 3,452 combinations (out of 259,560, see SI) that successfully create other elements. We believe that "Little Alchemy 2" is a quintessential game of exploration because players do not play for rewards but instead are intrinsically motivated to explore the game tree and create new elements. It offers a rich and semantically meaningful structure, which probes humans' intuitions about the combinability of its elements. Similar games have been used as a paradigm to study artificial agents' commonsense knowledge when trained on natural language corpora[15].

Uncertainty-Guided Exploration

How can and should people explore element combinations in "Little Alchemy 2"? We compare two different strategies in terms of how well they describe players' behavior in the game: uncertainty-guided exploration and exploration as empowerment.

One class of heuristics is to use one's uncertainty about different options to guide one's exploration behavior. For example, one way to implement simple uncertainty-guided exploration is to assume an uncertainty bonus that encourages the sampling of options that have not been sampled frequently in the past. Models of human exploration using this type of uncertainty-guidance have been very prolific, describing behavior in simple multi-armed bandits[5], bandits with correlational structures[16], as well as real-world decision-making problems[17]. Uncertainty-guided exploration, therefore, constitutes a good candidate model to describe human exploration in more complex paradigms as well. In "Little Alchemy 2", an uncertainty-guided strategy would correspond to tracking how often one has used particular elements before and then using the elements more that have not been used frequently. It can be formalized as
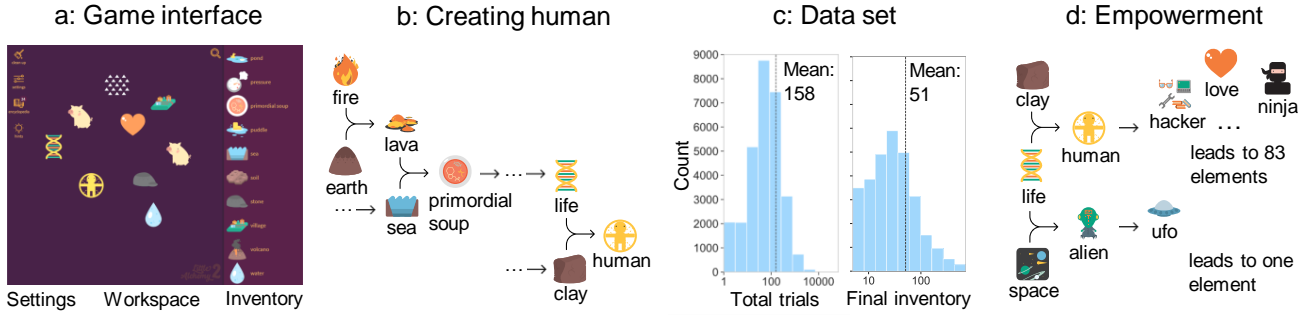
Figure 1. "Little Alchemy 2". **a:** Game interface. Players can use the workspace to combine two elements in an attempt to create new elements, which get added to the inventory and can be used for future combinations. **b:** Creating human. "Little Alchemy 2" is a richly structured game in which semantically meaningful combinations can create up to 720 unique elements. **c:** Data set of 29,493 players attempting a total of 4,691,033 combinations. Participants played for 158 trials and found 51 elements on average. **d:** Example of empowerment algorithm. Combining life with clay creates human, which can lead to 83 other elements in total, such as hacker, love, and ninja. Combining life with space creates alien, which can only lead to one other element, UFO. Thus, combining life with clay is the more empowering action.

$$U_e = \sqrt{\frac{\log(T)}{t_e}}, \tag{1}$$

where the uncertainty value U of each element e depends on the total number of trials T so far and the number of times t_e the element has been chosen in the past. The higher the uncertainty value of a combination (the sum of the individual values of the two respective elements), the higher the probability of choosing that combination.

Exploration as Empowerment

Other exploration strategies could also be at play in more complex scenarios such as "Little Alchemy 2". One such strategy is empowerment[18]. Exploration as empowerment centers around the idea of exploring options that enable the generation of as many more options as possible. This idea is also at the center of many examples of real-world exploration. For example, in the pharmaceutical sciences, researchers attempt to find new methods to produce vaccines even faster in the future. Children at play might also exhibit this kind of intrinsic drive to explore actions that train them to explore even more actions[9,19]. Empowerment differs from uncertainty-guided strategies, in that it focuses less on reducing uncertainty about the environment and more on the intrinsic drive towards a state of maximum influence[20]. It would not be easily possible to study exploration as empowerment in multi-armed bandits as the influence on future options cannot be changed during the experiment. Nonetheless, we believe that empowerment captures a quintessential component of intrinsic exploration such as in children's play and scientific investigations: the attempt to do things that enable one to do even more things in the future.

In the context of "Little Alchemy 2", empowerment translates into the players' intrinsic desire to create elements that offer many new successful combinations. We therefore call an element empowering the more distinct elements it can lead to by combination with all other possible elements in the game, provided that the true game tree is known. For example, the element human in combination with other elements leads to 83 new elements, while alien leads to only 1 new element (see Fig. 1d). Thus, the element human is more empowering than the element alien. The empowerment value of a combination is the empowerment value of the resulting element – so the number of elements it can create. In our example, the combination of space and life has a lower empowerment value (1) than clay and life (83). Without knowledge of the true game tree, empowerment requires both a semantic understanding of what elements could come out of a particular combination as well as an intuitive understanding of how potent, i.e. combinable, the resulting elements could be. Thus, studying empowerment requires a richly-structured and semantically-meaningful game tree such as the one afforded by "Little Alchemy 2".

Results

We look for signatures of uncertainty-guided exploration and empowerment in a data set of online players of the game "Little Alchemy 2" using statistical and cognitive modeling. Additionally, we change the structure of the game to investigate whether a simpler paradigm and a version without rich game semantics could lead to similar results. Further details can be found in the

Materials and Methods section, as well as in *SI Appendix*. In the first section of our *Results*, we present the online data set and some descriptive analyses. In the following two sections, we show that humans incorporate the empowerment value in their behavior and look at the performance of different models playing the game. Afterwards, we address people's intuitive semantic understanding of the game and how an approximation of this understanding can be integrated into the empowerment model. In the following section, we show that humans use a mixture of exploration as empowerment and uncertainty-guided exploration when playing the game. Finally, we extend our results by gathering two similar data sets from online experiments, omitting the semantic structure of the game in one of them.

Online Game Data

We collected data from anonymous online players of the game over a duration of three weeks, resulting in a data set of 29,493 players who tried over four million combinations. From each player, we know the whole course of their gameplay, that is the order of tried combinations, starting with the basic inventory of four elements. Players played for an average of 158 trials and discovered an average of 51 elements (see Fig. 1c; mean number of trials = 158.06; 95% CI 150.12, 165.99; mean number of elements = 50.91; 95% CI 50.03, 51.78). 563 players even played for longer than 1000 trials, with 16 of them playing over 10.000 trials. 3206 players managed to have an inventory with more than 100 elements; 9 players managed to find all possible 720 elements.

Drivers of Exploration Behavior

What strategies do humans use to explore the space of possible elements? Players immediately used elements that they had just created very frequently (see Fig. 2a). We therefore further analyzed what drove players to immediately use a new element after it had just been created. The idea of this analysis was that if people have a good intuition and care about empowerment, then they should immediately use empowering elements as soon as they have been created. This analysis showed that players had a preference to use an element immediately after creating it if the element had a higher empowerment value, i.e. the actual number of elements it could lead to. We assessed the size of the effect by comparing their choices with a simulated random performance, which revealed that they differed meaningfully (human: Kendall's $r_\tau = .43$, $p < .001$, 95% CI .38, .47; random: $r_\tau = .23$, $p < .001$, 95% CI .18, .27). This suggests that people incorporate the empowerment value of the different elements in their decision to immediately use a newly-created element.

Another aspect of people's playing behavior is the point in time at which they stop playing the game. What motivates players to continue combining elements? We analyzed whether continuation of play is more influenced by the recent creation of successful combinations, or by the recent creation of empowering elements. We regressed the value of the previous two rounds (number of successes in the success model and sum of empowerment values in the empowerment model) onto the decision of continuation of the players in the current round. We found that the empowerment value of discovered elements had a positive effect on continuation of play ($\beta = 0.41$, $z = 44.08$, $p < .001$), while the success value did not ($\beta = -0.30$, $z = -32.76$, $p < .001$, see Fig. 2b). This means that players' decisions to continue the game were mostly influence by how empowering recently created elements were.
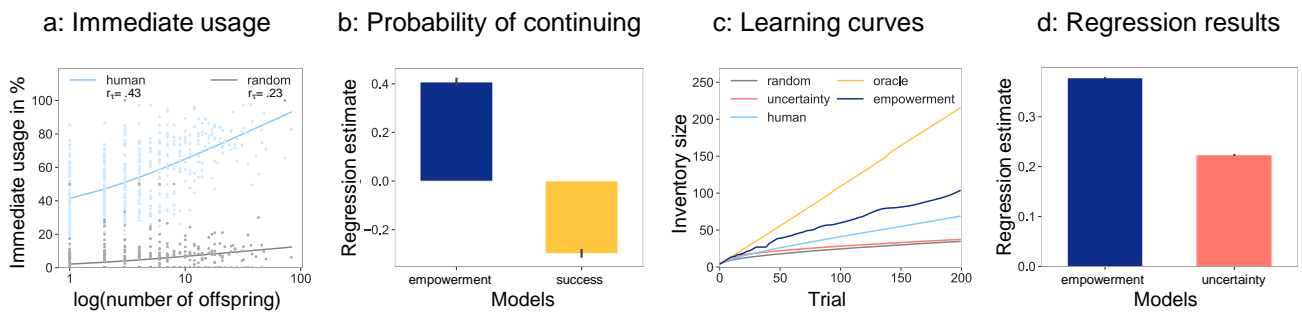

Figure 2. Empowerment results. **a:** Percentage of immediately used newly-discovered elements depending on their empowerment value - how many elements they can produce. Players were more likely to immediately use more empowering elements than would be expected under a random performance. **b:** Probability of continuing the game. While the empowerment value of recently discovered elements had a positive influence on participants' probability of continuing the game, the success of combinations had not. **c:** Performance of different models when playing "Little Alchemy 2". The uncertainty model performed marginally better than chance, while the empowerment model performed better than humans. The oracle model indicates the performance of an optimal agent. **d:** Model comparison. A combination of empowerment and uncertainty described human behavior in "Little Alchemy 2" best.

Model performance in playing the game

We assessed the performance of different models by letting them play the game from the beginning. We tested the performance of the empowerment approach by creating a model based on the empowerment values of the underlying game tree and compared this model to a random choice, an oracle, and an uncertainty-based exploration model.

The random model picks the elements of the next combination randomly from the current inventory. The oracle model knows the actual game tree and chooses combinations that always result in the discovery of a new element, thus simulating the behavior of a perfect agent. The uncertainty model picks elements based on how often they have been used so far (see Equation 1). The more often an element has been chosen, the less likely it is to be chosen again. The empowerment model bases its decisions on the empowerment value of the possible combinations.

The values of the latter two models were converted into probabilities using a softmax function before a combination was selected according to these probabilities. Each model also had a perfect memory, i.e. they never tried past combinations again. We ran each model 1,000 times over 200 trials. In Figure 2c, we plotted the average inventory size over time while also comparing to human players. The oracle model and the empowerment model outperformed human players. This was expected because both of these models knew the underlying game tree, while people did not. The uncertainty-based and the random model performed worse than humans. Since human performance was between the uncertainty and the empowerment model, it is conceivable that players were using a mixed strategy, similar to other theories of human learning and decision making[6, 21].

Approximating the intuitive semantics of the game

We believe that players have an intuitive understanding of which elements are combinable and empowering and which are not, which we operationalized in our empowerment model. The empowerment model must be based on a game tree to calculate the values of the different combinations. As we wanted to compare how the different models describe human behavior, we had to decide on a reasonable semantic basis for the empowerment model. Since humans do not know the true game tree, we did not match the true underlying game tree with players' decisions. Instead, we had to capture people's intuitive understanding of which elements can be combined and which cannot. Clearly, it would not be feasible to ask players about their intuitions about all possible 720×720 element combinations (see also SI). Thus, we decided to approximate human intuitions by approximating their semantic understanding using neural networks trained on parts of the game tree. We used a word representation model of vector embeddings pre-trained on a large English language corpus of Wikipedia articles[22]. Similar models have been used to model human judgement and decision making in other domains[23]. The elements' word vectors were used as inputs to two feedforward neural networks. The first model was a *link prediction model*, which predicted which element combinations were likely to succeed. The second model was an *element prediction model*, which assigned probabilities to each of the 720 possible elements, stating how likely the respective element was to result from the given combination. Both neural networks were trained on sub-parts of the true game tree, by dividing the possible 259,560 combinations into a training, validation, and test data set. We used 10-fold cross-validation such that each element combination was part of a test set at least once. For all further analysis, we only used the predicted probabilities of combinations that were not part of the training set. These probabilities were used to create new empowerment values for the different combinations by multiplying their success probability with the outcome probabilities of all elements times their specific empowerment values (see SI for more details). Thus, combinations that had a high probability to succeed and a high probability to result in an element that can create more elements in the future had a high empowerment value, predicted by the two models in unison. We validated this approach extensively by matching the models' predictions with people's intuitions in an additional experiment reported in the SI.

Regression analysis

We compared how well the different models described the actual behavior of all players. Because players' inventory grows over time, it becomes difficult to compare across all possible choices. Therefore, we decided to use a simpler method to compare the different models' predictions, which was to create a data set comparing the value of a combination chosen by a player according to the current model with the value of a randomly-sampled combination that the player could have chosen but did not. We then fed the differences of these two value predictions into a logistic mixed-effects regression[6], allowing us to regress model predictions onto participants' choices (see SI for further details including multiple recovery analyses).

We compared several models of players' gameplay while also controlling for the number of trials different players played. We found that the model best describing choices was a combinations of empowerment ($\beta = 0.38$, $z = 368.86$, $p < .001$) and uncertainty-guided ($\beta = 0.22$, $z = 204.64$, $p < .001$) exploration (see Fig. 2d; see SI for full model comparison). Importantly, the effect of empowerment was larger than the effect of uncertainty-guided exploration ($\beta = 0.118$, $z = 85.95$, $p < .001$). Although there was a positive effect of uncertainty-guided exploration, our model recovery results showed that we cannot fully distinguish between uncertainty-guided and random exploration. Thus, participants were driven mostly by the attempt to create elements that empowered them to create even more elements.

The importance of semantic structure

We argued that richly-structured environments should be used to study more sophisticated exploration strategies than are typically found in simpler paradigms such as multi-armed bandit tasks. In particular, we believe that "Little Alchemy 2" is richer than multi-armed bandits on two axes. First, multi-armed bandits are flat with only very few options, while our game has many possible nodes and ways in which an exploring player can go. Second, multi-armed bandits are also blank because they do not assume any prior knowledge, while our game requires a rich semantic understanding of how different elements could be combined. We hypothesized that people explore by trying to empower themselves in the game not only because the game contains a rich game tree but also because of the rich semantics underlying the tree. To further probe this hypothesis, we checked if stripping away the semantics of the game changed players' exploration strategies. For that purpose, we created our own two versions of the game which we termed "Tiny Alchemy" and "Tiny Pixels". Both versions are based on the game tree of "Little Alchemy 1", the predecessor of "Little Alchemy 2". The game "Tiny Alchemy" contains 540 elements and has a similarly rich semantic structure as the original game. The game "Tiny Pixels" has the same underlying game tree but we eliminated the semantics by randomly repositioning the pixels in the pictures of the elements and adding uninformative names usually used in memory tasks (see Fig. 4). With the semantic structure removed, the game is still rich (because it is not flat) but we expected it not to be rich in the right way (because it was now blank). We collected data for both games on Amazon Mechanical Turk, gathering 97 participants for "Tiny Alchemy" and 98 participants for "Tiny Pixels". The resulting data set allowed us to assess how much our finding of exploration as empowerment was indeed driven by the semantics of the game.
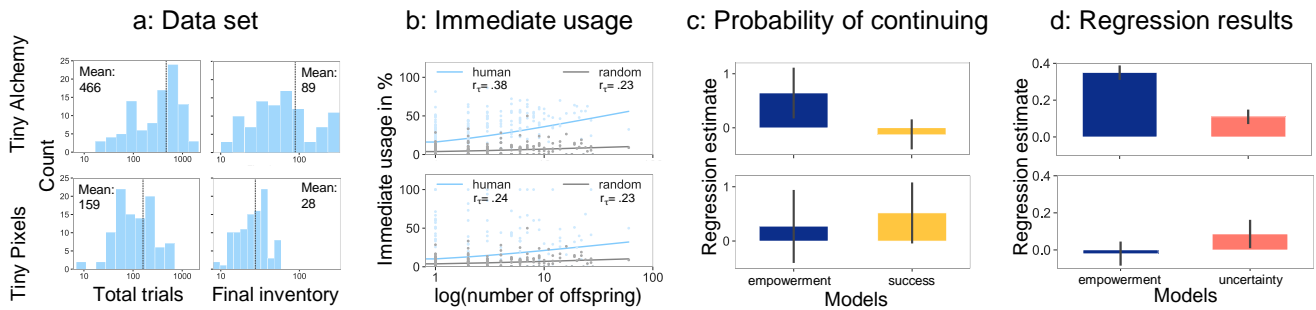

Figure 3. Importance of richly-structured game semantics. **a:** In "Tiny Alchemy", 97 participants played for 466 trials and created 89 elements on average. In "Tiny Pixels", 98 participants played for 89 trials and created 28 elements on average. **b:** Whereas participants were more likely to immediately use empowering elements in "Tiny Alchemy", this was not the case for participants in "Tiny Pixels". **c:** While the empowerment value of recently discovered combinations had an influence on participants' continuation behavior in "Tiny Alchemy", this was not the case in "Tiny Pixels". **d:** Model comparison. Whereas empowerment and uncertainty both significantly related to participants' choices in "Tiny Alchemy", only uncertainty-guided exploration –and not empowerment– did for "Tiny Pixels".

Behavioral differences

We first investigated simple behavioral differences between the two games under the assumption that "Tiny Pixels" would be more difficult for players than "Tiny Alchemy". We found that players of the game "Tiny Alchemy" played on average longer ($t(193) = 5.77$, $p < .001$, $d = 1.03$) and discovered more elements ($t(193) = 7.21$, $p < .001$, $d = .83$) than players of the game "Tiny Pixels" (see Fig. 3a; "Tiny Alchemy": mean number of trials = 465.55; 95% CI 385.39, 545.70; mean number of elements = 89.07; 95% CI 72.19, 105.95; "Tiny Pixels": mean number of trials = 81.0; 95% CI 61.89, 100.11; mean number of elements = 19.06; 95% CI 17.17, 20.95))).

Next, we compared the frequency of immediately using a newly-created element based on its empowerment value - the ability to create new elements later on. In "Tiny Alchemy", players were more likely to use empowering elements immediately than in "Tiny Pixels" (see Fig. 3b; "Tiny Alchemy" human: $r_\tau = .38$, $p < .001$, 95% CI .33, .43; "Tiny Pixels" human: $r_\tau = .24$, $p < .001$, 95% CI .20, .28; random: $r_\tau = .23$, $p < .001$, 95% CI .19, .28).

We again compared the influence of success and empowerment values of recent combinations on players' probability of continuing the game. In "Tiny Alchemy", players were more likely to continue the game when they had recently discovered empowering elements ($\beta = 0.64$, $z = 2.69$, $p < .01$), but were not when we just looked at the success of recent combinations ($\beta = -0.12$, $z = 0.14$, $p = 0.39$). In "Tiny Pixels", neither of the two models had a significant effect on participants' decision to continue the game (empowerment: $\beta = 0.27$, $z = 0.78$, $p = 0.43$, success: $\beta = 0.51$, $z = 1.80$, $p = 0.07$, see Fig. 3c).

Regression analysis for experimental data

We conducted a similar regression analysis as for the online data from before. However, we combined the "Tiny Alchemy" and "Tiny Pixels" data sets and included a variable indicating which version a player played. We again included the number of trials and the interactions of model predictions with the number of trials in our regression analysis to account for the unequal length of the data sets. For "Tiny Alchemy", players were best-explained by a combination between exploration as empowerment ($\beta = 0.322$, $z = -31.568$, $p < .001$) and uncertainty-guided exploration ($\beta = 0.100$, $z = 9.789$, $p < .001$), with the effect of empowerment being stronger than the effect of uncertainty-guided exploration ($\beta = 0.275$, $z = 19.910$, $p < .001$). For "Tiny Pixels", players' choices were only significantly influenced by uncertainty ($\beta = 0.080$, $z = 4.798$, $p < .001$) but not by empowerment ($\beta = 0.010$, $z = 0.610$, $p = 0.543$, see Fig. 3d), leading to a higher effect of uncertainty-guided exploration over empowerment ($\beta = -0.214$, $z = -9.315$, $p < .001$). These results further strengthen our idea that rich environments are necessary to study complex explorations strategies such as empowerment and that players' exploration looks more like what has been frequently found in traditional multi-armed bandits paradigms when the rich structure of the game is removed.

Discussion

We have studied human exploration in a richly-structured online game using a large data set of human playing behavior. Beyond just uncertainty-guided exploration, we were able to show that people take into account how empowering the outcomes of their actions are and that this intuition strongly guides their exploration behavior. Detailed computational modeling showed that these patterns could be captured quantitatively. Our results suggest that people use richly-structured and semantically meaningful exploration strategies in the game, resembling other strategies observed in the real world such a children's playing behavior or scientific methods of discovery.

Of course, our current empowerment model is just one formalization of exploration in richly-structured environments such as games. Even though we found our model to match well with both people's actual game playing behavior and their intuitions in a validation experiment, other strategies of exploration could also be assessed using our data. One such strategy is powerplay[24], which attempts to train one's model of the world as much as possible and would predict that players not only create elements to empower themselves but also to learn more about the game mechanics in general. Another strategy is goal-conditioned exploration[25], i.e. setting yourself goals to accomplish within the game. For example, in "Little Alchemy 2", having the goal of creating a solar cell would probably lead to a different exploration path through the game tree than the goal of creating chicken soup. There have also been several other studies on both the algorithmic[26–28] and behavioral[29,30] underpinnings of more sophisticated exploration strategies. In future investigations, we will attempt to compare more elaborate strategies of exploration using our data.

Relatedly, our current model does not incorporate any learning of the underlying structure and solely focuses on the exploration aspect of people's play. We believe that this is a good first step to understanding exploration as empowerment because people likely already have detailed intuitions about the different element pairs before the start of the game. Nonetheless, one of our future goals is to build models that update their intuitions while playing the game and thereby simulate people's learning progress.

Another concern is the fact that the underlying semantic structure of the game tree was designed by just one person, i.e. the creator of the game. Thus, one could argue that the game might not tell us much about people's general intuitions and exploration behavior. We do not believe this to be the case for two reasons. First, we were able to show that these intuitions are shared among the players within our validation study. Second, games are generally designed to be natural for people - meaning they have to be learnable and are calibrated to people's intuitive theories in the first place. Thus, we believe that games such as "Little Alchemy" can be used to reverse-engineer people's intuitive semantics and use them to model human exploration.

Finally, even though players participated in the online game "Little Alchemy 2" without any external rewards, participants of our experimental versions of the game, "Tiny Alchemy" and "Tiny Pixels", were rewarded for generating new elements. This means that participants exploring the game intrinsically behaved similarly to participants who participated in our online experiments for monetary rewards. We used the experimental versions of the game to establish that stripping away the semantics of the game changed participants exploration strategies after having established intrinsic exploration strategies using data from the online game already. Nonetheless, in future studies, we would like to further disentangle the effects of rewards on players' behavior by also removing the semantics of the online game.

Taken together, our results advance our understanding of human intrinsic exploration behavior and extend current research paradigms by using a large, complex, and richly-structured data set of an online game. One implication of our results could be that empowerment –or other more elaborate exploration strategies– may often drive people's decisions but are masked by the simple paradigms used in research on human exploration strategies. Perhaps more sophisticated strategies can simply not be found in easier paradigms or look like simpler strategies, such as uncertainty-guided exploration, when studied in reduced forms. Thus, we believe that our work demonstrates that using games as experimental paradigms can increase the complexity, robustness, and ecological validity of psychological research.

Conclusion

We investigated the exploration behavior of 29,493 players in the richly-structured online game "Little Alchemy 2". We found that players' behavior was mostly driven by the attempt to empower themselves, i.e. they created elements that can lead to more elements over time. Using two additional games, we replicated our results in a controlled setting and showed that participants resorted to simpler exploration strategies when the semantic structure of the game was removed. Our results point to the necessity to use more complex experimental paradigms to study elaborate strategies of human exploration. We hope that our findings and model are a first step towards empowering our own theories of human exploration.

Material and Methods

"Little Alchemy 2" Data Set

The "Little Alchemy 2" data set was gathered over a duration of three weeks from June 1st to June 21st, 2019 with the help of the game's developer. For all our analyses, we only included players who started to play the game within that time period and filtered out all repeated trials. This led to a data set containing 29,493 players who tried 4,691,033 combinations in total. All included players consented to their anonymized data being used for scientific purposes.

"Tiny Alchemy" and "Tiny Pixels"

For the experimental versions of the game, i.e. "Tiny Alchemy" and "Tiny Pixels" we recreated the game using the game tree of "Little Alchemy 1". Whereas players of "Tiny Alchemy" played the game with normal element pictures and names, "Tiny Pixels" used element pictures with randomly positioned pixels and unrecognizable yet distinct names (see Fig. 4). Players were paid $0.10 for every discovered element and could play for as long as they wanted but up to 2 hours. We recruited participants from Amazon's Mechanical Turk. For "Tiny Alchemy", we recruited 97 participants. For "Tiny Pixels", we recruited 98 participants. All experiments were approved by Harvard's Institutional Review Board.
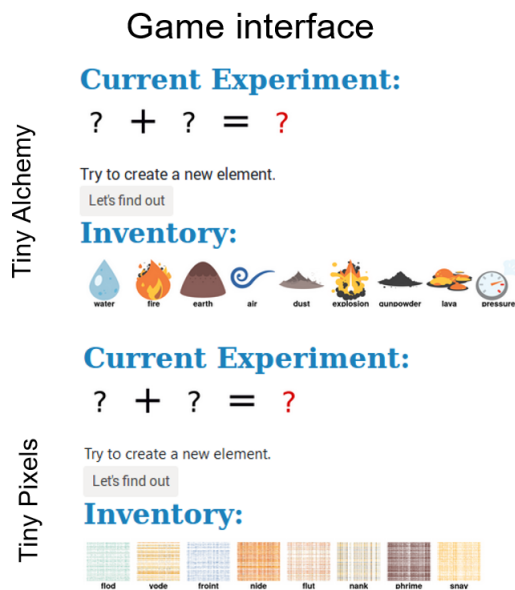

Figure 4. Experiment setup of "Tiny Alchemy" and 'Tiny Pixels". "Tiny Alchemy" is an experimental version of "Little Alchemy 2". "Tiny Pixels" is based on "Tiny Alchemy" but does not contain any semantic information.

Empowerment Model

The empowerment model is based on how many distinct elements an element can produce by combination with any other element. To calculate this value, we needed to simulate human intuitions about the game tree. This process consisted of four steps: Pre-processing the words by turning them into word vectors, prediction of the link probability, prediction of the resulting element, and the actual calculation step.

First, we used a pre-trained word representation model of word vector embeddings called fasttext[22]. Thereby, we got 300-dimensional word vectors for each of the elements. For each combination, we then concatenated the two vectors of the involved elements to use them as one combination vector.

Second, we used the vectors as input combinations for a link prediction model, which consisted of a fully connected neural net with one hidden layer. This model returned a link probability for the input combination vector. It was trained on sub-parts of the true game tree - all 259,560 possible combinations were split into a training, validation, and test data set. We used 10-fold cross-validation such that each element combination was part of a test set at least once. For our further analysis, we only used combinations' predicted probabilities that were not used in training. If a combination had a predicted value of higher than 50 %, it was classified as a link.

Third, we used the concatenated word vectors as inputs for the element prediction model, which was another fully connected neural network with two hidden layers. This model returned the probability of being the resulting element for each of the 720 possible elements, based on the cosine similarity of the word vector. As before, training was conducted using 10-fold cross-validation.

Fourth, the resulting values of each combination were merged into an empowerment value by multiplying the predicted probability of success with the sum of the probabilities of each element multiplied with the expected empowerment value of an element - the number of distinct elements resulting from combinations involving the element with a predicted success greater or equal to 0.5.

Acknowledgments

References

1. Schulz, E. & Gershman, S. J. The algorithmic architecture of exploration in the human brain. *Current opinion neurobiology* **55**, 7–14 (2019).

2. Wilson, R. C., Bonawitz, E., Costa, V. D. & Ebitz, R. B. Balancing exploration and exploitation with information and randomization. *Current Opinion Behavioral Sciences* **38**, 49–56 (2021).

3. Daw, N. D., O'doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).

4. Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A. & Cohen, J. D. Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal Experimental Psychology: General* **143**, 155–164 (2014).

5. Speekenbrink, M. & Konstantinidis, E. Uncertainty and exploration in a restless bandit problem. *Topics cognitive science* **7**, 351–367 (2015).

6. Gershman, S. J. Deconstructing the human algorithms for exploration. *Cognition* **173**, 34–42 (2018).

7. Gershman, S. Uncertainty and exploration. *Decision* **6**, 277–286 (2019).

8. Brändle, F., Binz, M. & Schulz, E. Exploration beyond bandits. *PsyArXiv* , (2021).

9. Chu, J. & Schulz, L. Exploratory play, rational action, and efficient search. *PsyArXiv* , (2020).

10. Gottlieb, J., Oudeyer, P.-Y., Lopes, M. & Baranes, A. Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends cognitive sciences* **17**, 585–593 (2013).

11. Payzan-LeNestour, E. & Bossaerts, P. Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Computational Biology* **7**, e1001048 (2011).

12. Knox, W. B., Otto, A. R., Stone, P. & Love, B. C. The nature of belief-directed exploratory choice in human decision-making. *Frontiers Psychology* **2:398**, 1–12 (2012).

13. Schulz, E., Wu, C. M., Ruggeri, A. & Meder, B. Searching for rewards like a child means less generalization and more directed exploration. *Psychological science* **30**, 1561–1572 (2019).

14. Little alchemy 2. https://play.google.com/store/apps/details?id=com.recloak.littlealchemy2 (2021).

15. Jiang, M. *et al.* Wordcraft: An environment for benchmarking commonsense agents. *arXiv preprint arXiv:2007.09185* , (2020).

16. Schulz, E., Franklin, N. T. & Gershman, S. J. Finding structure in multi-armed bandits. *Cognitive psychology* **119**, 101261 (2020).

17. Schulz, E. *et al.* Structured, uncertainty-driven exploration in real-world consumer choice. *Proceedings National Academy Sciences* **116**, 13903–13908 (2019).

18. Klyubin, A. S., Polani, D. & Nehaniv, C. L. All else being equal be empowered. In *European Conference on Artificial Life*, 744–753 (Springer, 2005).

19. Colantonio, J. & Bonawitz, E. Awesome play: Awe increases preschooler's exploration and discovery. *OSF Preprints* , (2018).

20. Salge, C., Glackin, C. & Polani, D. Empowerment–an introduction. In *Guided Self-Organization: Inception*, 67–114 (Springer, 2014).

21. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204–1215 (2011).

22. Joulin, A. *et al.* Fasttext. zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651* , (2016).

23. Bhatia, S. Associative judgment and vector space semantics. *Psychological review* **124**, 1 (2017).

24. Schmidhuber, J. Powerplay: Training an increasingly general problem solver by continually searching for the simplest still unsolvable problem. *Frontiers psychology* **4**, 313 (2013).

25. Nasiriany, S., Pong, V. H., Lin, S. & Levine, S. Planning with goal-conditioned policies. *arXiv preprint arXiv:1911.08453* , (2019).

26. Campero, A. *et al.* Learning with amigo: Adversarially motivated intrinsic goals. *arXiv preprint arXiv:2006.12122* , (2020).

27. Chitnis, R., Silver, T., Tenenbaum, J., Kaelbling, L. P. & Lozano-Perez, T. Glib: Efficient exploration for relational model-based reinforcement learning via goal-literal babbling. *arXiv preprint arXiv:2001.08299* , (2020).

28. Pathak, D., Gandhi, D. & Gupta, A. Self-supervised exploration via disagreement. In *International conference on machine learning*, , 5062–5071 (PMLR, 2019).

29. Gottlieb, J. & Oudeyer, P.-Y. Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience* **19**, 758–770 (2018).

30. Chu, J. & Schulz, L. E. Play, curiosity, and cognition. *Annual Review Developmental Psychology* **2**, 317–343 (2020).

Supporting Information

The "Little Alchemy 2" dataset

The original "Little Alchemy 2" data set was collected on the game's online platform during a period of three weeks from June 1st to June 21st 2019. The game – and therefore also our dataset – has some additional features that were not mentioned in the main text. First, a combination of two elements in "Little Alchemy 2" can –in rare cases– create more than one element (up to three). In this case, all elements get added to the player's inventory. Second, elements can be final, i.e. they cannot create other elements, and are not added to the inventory, which is indicated to the player. There is also an encyclopedia, where players can look up all elements they have created so far. Third, elements can be depleted, i.e. the player found all successful combinations with that element, and thereby eliminated it from the usable inventory (but it can still be found in the encyclopedia). Finally, there exist some elements that are added to the inventory once a player fulfills additional requirements. For example, "metal" is added once the player's inventory consists of 50 elements. There exist nine of these "unlockable" elements.

The initial and unfiltered dataset consisted of 32,177 players and 6,692,132 trials in total. We did not have access to the full play history for some of these players because they had started playing the game before data collection commenced. We removed these players from all of our analyses. We also removed all duplicate trials because it was unclear why players repeated an element combination. The final data set consisted of 29,493 players and 4,691,033 trials in total.

Tiny Datasets

In the "Tiny Alchemy" and "Tiny Pixels" datasets, we eliminated the complications of the original dataset by not telling the players if an element was final or depleted and leaving the elements in the usable inventory. We also did not implement "unlockable" elements. The experiment was based on the game tree of "Little Alchemy 1", the predecessor of "Little Alchemy 2". To simplify the game further, we changed the game tree so that only one resulting element per combination was possible, by omitting all other results from the combination if existing in the original game tree. Whereas players of "Tiny Alchemy" played the game with normal element pictures and names, "Tiny Pixels" used randomly-sampled, pixelated element pictures and unrecognizable yet distinct names.

The game interface lets participants always select two elements for an "experiment". Once participants wanted to submit a combination, they could click on a "Let's find out" button. If the experiment created a new element, then this element was

added to their inventory. If the experiment did not create a new element, a red cross appeared as outcome, and participants were told that the experiment was unsuccessful.

In both versions, players were paid $0.10 for every discovered element and could play for as long as they wanted but up to 2 hours. We recruited participants from Amazon's Mechanical Turk. For "Tiny Alchemy", we recruited 97 participants. For "Tiny Pixels", we recruited 98 participants. All experiments were approved by Harvard's Institutional Review Board. Participants were required to be from North America, as well as to have a past HIT completion rate of at least 99% with a minimum of 100 past completed HITs. Participants were only allowed to participate in one of our experiments.

Approximating the game tree

For our model comparison, we approximated the intuitions of the players about the different elements. As explained in the method section of the main article, we used a combination of pre-learned word vectors and two feedforward neural networks, which were trained on the true game tree, to determine the probability that the combination of two elements is successful and the probability for each element to be the result.

We used fasttext –a pre-trained word representation model of word vector embeddings– to get a word vector for every element. We chose a version of fasttext that was pre-trained on a large English language corpus of Wikipedia articles. We used fasttext because one can create word vectors of words that were not seen during training by using subword information. This was important because the elements of "Little Alchemy 2" can be very specific – like double rainbow – and were probably not part of the training set of the vector embeddings. Fasttext returns default vectors of dimension 300, which can be reduced to lower dimensions. We compared a model with 300 with a model with 100-dimensional word vectors.

As each of the elements can be combined with every other element including itself, but the order of the elements in the combination is not important, this led to 259.560 combinations in total. These combinations were used as the inputs to the two neural networks.

To approximate human intuition, we only wanted to use the predictions for combinations that were not used in the training set. Thus, we split all combinations into three datasets: training, validation and test. We used 10-fold cross-validation such that each combination was part of a test set at least once.
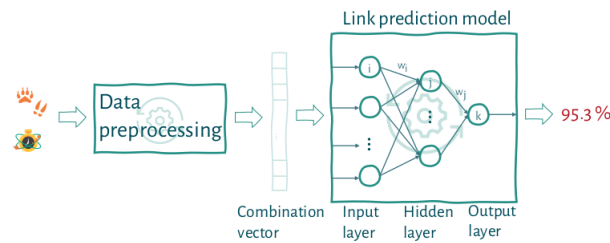

Figure 5. Structure of the link prediction model. Element combinations of the initial dataset were pre-processed, resulting in a combination vector. This vector was used as the input for a feedforward neural network with one hidden layer. The network returned a link probability, i.e. the probability that the element combination is successful and creates an element as output.

The first neural network is the link prediction model, which predicts whether or not a combination is successful in creating a new element (see Figure 5). As the dataset was unbalanced –only 1.27% of all combinations successfully create an element– we decided to use random oversampling for our training data. After the oversampling, the word vectors of the two elements in each combination were concatenated in both orders to use them as input for the neural network. Additionally, the data was normalized (based on the training dataset). The feedforward neural network consisted of an input layer of 600 or 200 neurons, dependent on the model (2 times the input vector dimensions, as the concatenated vectors each consisted of two elements), a hidden layer of 16 neurons (with a rectified linear unit as activation function) and an output layer of 1 neuron (with a sigmoid function as activation function). It was trained over 100 epochs with a step size of 30 per epoch and a batch size of 32. We used a binary cross-entropy loss function and the Adam optimization algorithm to adjust the weights (with a learning rate of 0.001). We compared two models, one using 100 dimensional vectors and one using 300 dimensional vectors. We trained both models until they converged. The model with vector size of 300 dimensions performed better than the model with a vector size of 100 dimensions (see Table S1). We therefore chose the model using 300 dimensional vectors as the final link prediction model. While the model performed well on the loss (see Figure 6), we observed a high false positive rate on the actual data (precision of 6.32%). However, we think that this approximates human intuition well, as players tried out many unsuccessful combinations (only 33.62% of all combination trials done by humans were successful), implying that they overestimated the probability of links.
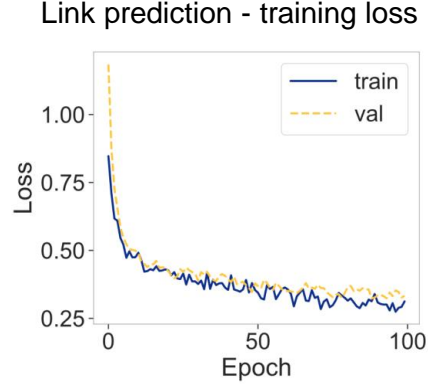
Link prediction - training loss

Figure 6. Loss of the link prediction model with 300 dimensions in the 10 cross validation runs over 100 epochs.

Table 1. Averaged results for the link prediction model after 10 cross validation runs. (FN - false negatives, FP - false positives, TP - true positives, TN - true negatives, AUC - area under the curve)

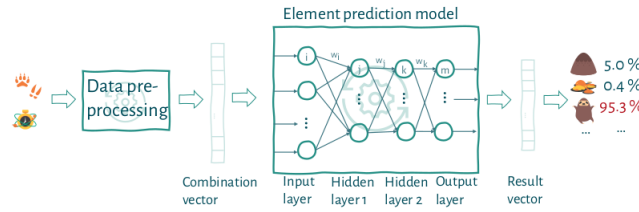| Vector	Loss	FN	FP	TP	TN	Precision	Recall	AUC
300	0.329	147	7073	477	44215	0.0632	0.7644	0.9006
100	0.396	149	8732	513	42518	0.0555	0.7758	0.8826

Figure 7. Structure of the element prediction model. Element combinations of the initial dataset were pre-processed, resulting in a combination vector. This vector was used as input for a feedforward neural network with two hidden layers. The network then returned a predicted result vector which was transformed into outcome probabilities for each of the 720 elements.

The second neural network is the element prediction model, which predicts which element is likely to result from a successful combination. It learned a mapping of element combinations represented as vectors to a resulting element vector. This resulting vector was then used to predict the probability for each element to be the outcome of an input combination (see Figure 7). In contrast to the previous model, we eliminated the unsuccessful combinations from the training and validation data. This reduced the two data sets as only 3452 successful combinations exist. Here, no oversampling was necessary. The feedforward neural network consisted of an input layer of 600 or 200 neurons (dependent on the model) two hidden layers with 1034 neurons each (with a rectified linear unit as activation function) and an output layer of 300 neurons (with a linear function as activation function). We trained the network over 15 epochs with a batch size of 32 and used the cosine similarity between the predicted and actual element's embedding as the loss function. As before, we employed Adam optimization with a learning rate of 0.001 to adjust the weights. As the model did not yet learn the probabilities but actual result vectors, we compared the result vector of an input combination with each possible element vector to get a similarity estimate, which was then transformed into a probability using a soft-maximization function. This probability stated for each element how likely it was to be the result of the given input combination. To get a measure of the network's quality, we ranked the similarities to see how close the generated result vector came to predicting the true result (see Figure 8).

As the model with vector size 300 outperformed the model with vector size 100 in both neural networks, we decided to use the model with vector size 300 for our further analyses. Once the two neural networks with their corresponding cross-validation rounds were completed, the results for each possible input combination were merged by getting the mean value.
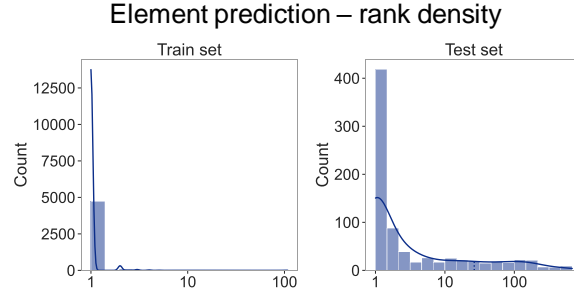
Figure 8. Predicted ranks of the element prediction model with 300 dimensions for the training and the test set as histogram and as density estimation (blue line).

Empowerment Model

The empowerment model is based on the concept that the more distinct elements a combination's resulting element can lead to, the more empowering it is. We therefore needed to base the model on a game tree, which can either be the true game tree – here the empowerment value of the combination is just the absolute number of distinct elements the resulting element can create – or our recreated version. To calculate the value from our recreated version, we used the predictions from the two learned neural networks, according to this formula.

$$E(e_{c_{A,B}}) = P(\text{link}_{c_{A,B}}) \times \sum_{i=0}^{720} P(\text{result}_{c_{A,B}} = i) \times E(e_i) \tag{2}$$

The expected empowerment value $E(e_{c_{A,B}})$ of the combination $c_{A,B}$ was composed of three parts: The link probability $P(\text{link}_{c_{A,B}})$, which was calculated by the first neural network, and the individual probabilities of each element being the result of the combination $P(\text{result}_{c_{A,B}} = i)$ multiplied with the individual empowerment value of the potential result. The individual empowerment values were estimated by looking at how many successful combinations ($P(\text{link}_{c_{i,j}}) \geq 0.5$) the elements were part of and how many distinct elements they were expected to produce (by assuming the element with the highest probability of each combination to be the resulting element $P(\text{result}_{c_{i,j}})$). At each trial, the empowerment value ($E(e_{c_{A,B}})$) of each available combination was transformed into a probability of choosing this particular combination at that time point by using a soft-maximization function with a temperature value of 0.1 (see section "Learning Curves"). The distribution of empowerment values in the true game tree and in our recreated version can be seen in Fig. 9. There was a substantial correlation between the model's predicted empowerment values and the true empowerment value of $r = 0.83$, $p < .001$.
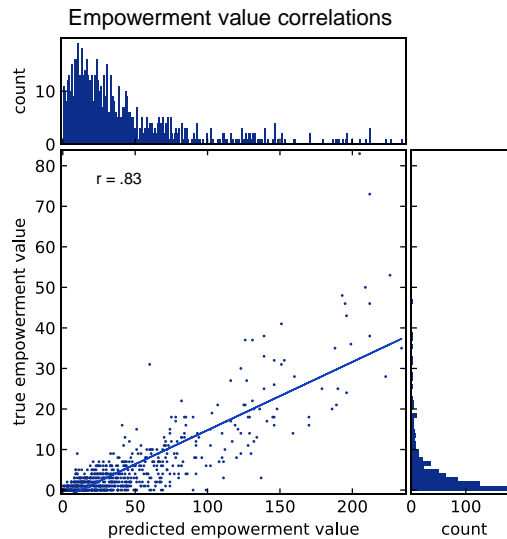

Figure 9. Correlation between the predicted empowerment values according to the approximated game tree and the true empowerment values according to the true game tree of "Little Alchemy 2".

Uncertainty model

The uncertainty model picks combinations that have not been used in the past more often.

$$U_e = \sqrt{\frac{\log(T)}{t_e}}$$

It assigns values to each element (e) based on the total number of trials so far (T) and the number of times the element has been chosen (t_e). The total value of a combination is the sum of the individual values of the two elements. These values are again transformed into probabilities at each timestep by using a soft-maximization function with a temperature value of 0.1 (see section "Learning Curves").

Baseline model

The baseline model picks combinations at random. At each time point, it chooses a random combination of two elements from the current inventory.

Success-Only or Empowerment?

Since our model formalizing exploration as empowerment consists of two components, a link prediction, and an element prediction component, one might ask if the link prediction model alone might already be enough to explain human behavior in "Little Alchemy 2". This would correspond to players only caring about whether or not a combination can successfully create new elements. To test this assumption, we first assessed whether or not there was more to people's choices of element combinations. We, therefore, first regressed the predictions of only the link prediction model onto players' choices. This showed a significantly positive effect of the link prediction component ($\beta = 0.379$, $p < .001$). Yet adding the predictions of the empowerment model to the regression model further improved the overall model fit ($\chi^2(2) = 6314.6$, $p < .001$, $\beta = 0.113$, $p < .001$). This served as a first indicator that players used more than just a "success-only" model to solve the task at hand. We also repeated this analysis only for players who played for at least 100 trials and only for their first 100 attempts. We did this for two reasons. First, we wanted to get a cleaner estimate of the two effects without having to correct for the different number of trials per player. Secondly, we wanted to remove the players who only tried out a few combinations and then stopped playing because the two models were less distinguishable for these players. Regressing both the link prediction model's and the full empowerment model's predictions onto participants' choices revealed a small effect of the link prediction model ($\beta = 0.015$, $p < .001$) and a large effect of the empowerment model ($\beta = 0.160$, $p < .001$). Therefore, we concluded that players were not only driven by the attempt to create new elements successfully but also tried to create elements to empower themselves.

Immediate Usage of Elements

To analyze the immediate usage of newly created elements, we excluded all final elements, as players could not use them after creation. Additionally, we also filtered out the four basic elements – water, earth, fire, air – as players had them in their inventory from the start. We simulated a dataset of 1000 players playing 1000 trials each according to our random model as a comparison. In "Tiny Alchemy" and "Tiny Pixels", we did not tell participants if an element was final, so they were still able to use them after creation. Therefore, we only needed to exclude the first four elements. As "Tiny Pixels" and "Tiny Alchemy" are based on the same game tree, we only created one random dataset for comparison. We used Kendall's τ throughout for the correlation analysis.

Probability of Continuing

We analyzed the probability of players continuing the game based on their previous two trials. We considered a success and an empowerment model. To calculate the value of the success model, we assigned 1 to trials in which the player discovered a new element, and 0 to trials in which the player did not. We then averaged the values of the previous two trials to get the success value for the current trial and normalized the resulting vector before entering it into the regression. For the empowerment model, we assigned 0 to trials in which the player did not discover an element, and the empowerment value of the element according to the true game tree, if the player did. We again averaged the values of the previous two rounds and normalized the vector to get the value of the current round. We regressed these values on the player's decision to continue the game: 1 if the player continued the game after this trial and 0 if the player stopped.

Learning Curves

The learning curves of the different models were created by letting them run over 200 steps for 1000 times and averaging the inventory sizes of all the runs. All models used a full memory, so they only picked combinations that had not been picked before. In Figure 2c, the oracle and the empowerment model were based on the true game tree. The oracle model

deterministically chose a successful combination with a resulting element that is not yet in the inventory. The baseline model chose the next combination randomly. The empowerment model and the uncertainty model both transformed their utility values u into probabilities by using a softmax function:

$$P_{c_i} = \frac{e^{\frac{u_{c_i}}{T}}}{\sum_{j=1}^{k} e^{\frac{u_{c_j}}{T}}}$$

The probability of choosing a certain combination c_i depends on the number of available combinations k and a temperature parameter T. In all our simulations, we set the temperature value to 0.1.

Regression Analysis

We performed a regression analysis to evaluate how the different models describe human behavior. As the inventory of players grows over time, it becomes difficult to compare all possible choices. Therefore, we reduced the problem to a two-armed bandit, by randomly assigning the value of the actually chosen combination to one arm and the value of a randomly sampled combination from the inventory of the player at that time – a combination that the player could have chosen, but did not – to the other arm. We then got the value difference between the first and the second arm and fed it into a logistic mixed-effects regression. If players chose the first arm, the decision label was set to 1, if not it was set to 0. Using this approach, we performed a bottom-up model comparison. The full results of this analysis can be found in the section "Full Regression Results".

Recovery Analysis

We simulated the data as described in the "Learning Curves" paragraph for the four models: random, uncertainty, empowerment, and success-only (for more information on the success-only model, see section "Success-Only or Empowerment?"). As for the learning curves, we let the models run 1000 times over 200 time steps. For the regression analysis, we used the same method as for the analysis of human data but added the values of the success-only model to the full model. Additionally, as all runs had the same number of time steps, we excluded the number of trials from the regression analysis. In the dataset created by the empowerment model (see Fig. 10a), empowerment had the largest effect ($\beta = 3.10$, $p < 0.001$), in comparison to uncertainty ($\beta = 0.83$, $p < 0.001$) and success-only ($\beta = 0.54$, $p < 0.001$). In the dataset created by the success-only model (see Fig. 10b), success-only had the largest effect ($\beta = 10.04$, $p < 0.001$), in comparison to empowerment ($\beta = -0.14$, $p < 0.001$) and uncertainty ($\beta = 1.32$, $p < 0.001$). In the dataset created by the uncertainty model (see Fig. 10c), uncertainty had the largest effect ($\beta = 22.61$, $p < 0.001$), in comparison to empowerment ($\beta = -0.10$, $p < 0.001$) and success-only ($\beta = 0.02$, $p < 0.05$). Here, we added noise to the simulated uncertainty values (from a Gaussian distribution with mean 0 and variance 0.05), as the uncertainty model achieved a perfect separation. In the randomly-created dataset, there still was an effect of uncertainty ($\beta = 0.46$, $p < 0.001$), but almost no effects of empowerment ($\beta = -0.03$, $p < 0.001$) and success-only ($\beta = -0.02$, $p < 0.05$) (See Fig. 10d). This could be the case as the inventory of the model and thereby the proportion of unexplored elements grows over time. Therefore, our current procedure may overestimate the effect of uncertainty.
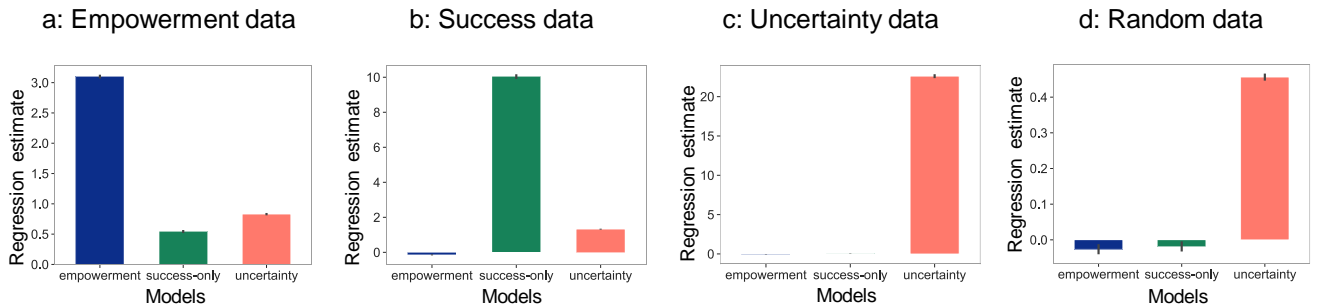

Figure 10. Recovery analysis. **a:** For data generated from the empowerment model, empowerment had the biggest effect. **b:** For data generated from the success-only model, success-only had the biggest effect. **c:** For data generated from the uncertainty model, uncertainty had the biggest effect. **d:** For data generated from the random model, uncertainty had the biggest effect.

Model Validation Experiments

We recruited 103 participants (mean age=37.1, SD=8.00, 86 females) from Amazon Mechanical Turk to take part in an experiment to further validate the empowerment model's predictions. Participants received a fixed payment of$2 for their

participation and the experiment took 25.8 minutes on average. The experiment was divided into three parts. In the first part, participants played the game "Tiny Alchemy" for 20 trials to get acquainted with the mechanics of the game. In the second part, participants always saw two random elements and had to indicate, on a scale from 1 (Definitely Not) to 7 (Absolutely), how likely they thought the two elements were to create a new element. No participant ever saw a combination that they had tried out during the first part of the experiment. We chose a subset of 1000 possible combinations with a wide range of predicted probabilities of success according to our link prediction model, and aimed for a balance between successful and unsuccessful combinations, according to the true game tree. Complete balancing was not possible as some ranges of the predicted probabilities did not include many successful combinations. For each participant we randomly sampled 20 out of these 1000 combinations. After having rated the possible success of a combination, four elements appeared underneath the combination and participants had to choose, under the assumption that the two current elements (the one that they just had rated) possibly would create an element, which of the four shown elements they thought it could be. The second part also contained 20 trials in total. The four possible elements consisted of the element with the highest probability predicted according to our approximated game tree (rank 1), one with a high probability (randomly sampled from rank 2-30), one with a medium probability (randomly sampled from rank 31-150) and one with a low probability (randomly sampled from rank 151-540). Finally, in the last part, participants had to rate different elements based on how useful they thought they were in the game, where we told them before that an element could be more useful if it could be successfully combined with multiple other elements, again on a scale from 1 (Not at all useful) to 7 (Very useful) and over 20 trials. For each participant, 20 elements were sampled from all possible 540 elements so that they cover a wide range of predicted empowerment values.

We were first interested in checking if our link prediction model captured parts of people's intuitions about the possible success of element combinations. We, therefore, regressed the *z*-transformed predicted probabilities onto participants' ratings of how likely they thought the two elements would create a new element in a mixed-effects regression while also entering a random intercept over participants. This analysis revealed that there was a significant fixed-effect of our model's predicted probability onto participants' ratings ($\beta = 0.132$, $t = 3.784$, $p < .001$) and that our model explained a significant amount of variance in participants' ratings (pseudo-r^2=0.336). Notably, even for cases in which the actual element combination did not produce an element according to the true underlying game tree, our model still matched well with participants' ratings ($\beta = 0.310$, $t = 6.334$, $p < .001$, pseudo-r^2=0.414). Thus, our link prediction model corresponds well with people's intuitions about the probability of success for a given combination.

Next, we analyzed which out of 4 candidate elements participants chose as the element they thought would likely result from the given combination. The average probability of the elements participants chose was much higher than for all other provided elements (0.41 vs. 0.19, $t(204) = 12.82$, $p < .001$, $d = 1.79$). This was even true for element combinations that actually did not successfully produce an element according to the rules of the game (0.39 vs. 0.21, $t(204) = 9.61$, $p < .001$, $d = 1.34$). We also regressed the (z-transformed) predicted probabilities onto whether or not a participant chose an element using a mixed-effects logistic regression. This showed a strong effect of the predicted probabilities onto participants choosing an element ($\beta = 0.882$, $z = 10.16$, $p < .001$, pseudo-r^2=0.270). This effect again remained even when controlling for combinations that did not lead to elements within the game ($\beta = 0.711$, $z = 12.173$, $p < .001$, pseudo-r^2=0.172). Thus, our element prediction model corresponds well with people's intuition about which element could result from a given combination.

Finally, we looked at how participants' ratings of how useful a particular element would be related to our model's prediction of empowerment values. For this, we regressed both the predicted and the actual empowerment values onto participants' ratings in a mixed-effects regression while adding a random intercept over participants. Since both the predictions and the true empowerment values were not normally-distributed (most elements only have a low empowerment value), we logarithmized the predictors and afterwards z-transformed them before performing the regression. The results of this analysis showed that there was a significant effect of the predicted empowerment values ($\beta = 0.146$, $t = 2.70$, $p < .001$) but not for the true empowerment values ($\beta = -0.027$, $t = -0.49$, $p = .617$) onto participants' ratings.

These results indicate that both the link prediction and the empowerment components of our model related meaningfully to participants' judgements about related properties of the game. Thus, we conclude that our model's components correspond well with people's intuitions about the game.

Full Regression Results

We report the full regression results for both "Little Alchemy 2" as well as the two lab versions, "Tiny Alchemy" and "Tiny Pixels".

Table 2. Full regression results for model comparison run on the "Little Alchemy 2" data set.

	Estimate	Std. Error	z value	Pr($>$|z|)
Uncertainty	0.22	0.00	204.64	0.00
Empowerment	0.38	0.00	368.86	0.00
Trial	-0.00	0.00	-0.02	0.98
Uncertainty $\times$ Trial	-0.61	0.00	-211.53	0.00
Empowerment $\times$ Trial	-0.00	0.00	-1.29	0.20

Table 3. Full regression results for model comparison run on "Tiny Alchemy" and "Tiny Pixels".

	Estimate	Std. Error	z value	Pr($>$|z|)
Uncertainty	0.09	0.01	8.86	0.00
Empowerment	0.30	0.01	29.68	0.00
Trial	0.00	0.01	0.55	0.58
Alchemy	0.00	0.01	0.12	0.90
Pixels	0.00	0.02	0.27	0.78
Uncertainty $\times$ Trial	0.04	0.01	4.70	0.00
Empowerment $\times$ Trial	0.11	0.01	10.46	0.00
Uncertainty $\times$ Pixels	-0.00	0.02	-0.19	0.85
Empowerment $\times$ Pixels	-0.35	0.02	-17.66	0.00