Using Games to Understand the Mind

Kelsey Allen[1,+], Franziska Brändle[2,+], Matthew Botvinick[1], Judith E. Fan[3], Samuel J. Gershman[4], Alison Gopnik[5], Thomas L. Griffiths[6], Joshua K. Hartshorne[7], Tobias U. Hauser[8,9,12], Mark K. Ho[6], Joshua R. de Leeuw[10], Wei Ji Ma[11], Kou Murayama[12], Jonathan D. Nelson[13], Bas van Opheusden[6], Thomas Pouncy[4], Janet Rafner[14], Iyad Rahwan[15], Robb B. Rutledge[16], Jacob Sherson[14], Özgür Şimşek[17], Hugo Spiers[8], Christopher Summerfield[18], Mirko Thalmann[2], Natalia Vélez[4], Andrew J. Watrous[19], Joshua B. Tenenbaum[20], and Eric Schulz[2,*]

[1]DeepMind, London, UK
[2]Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[3]Stanford University, Stanford, USA
[4]Harvard University, Cambridge, USA
[5]University of California, Berkeley, Berkeley, USA
[6]Princeton University, Princeton, USA
[7]Boston College, Boston, USA
[8]University College London, London, United Kingdom
[9]Max Planck UCL Centre for Computational Psychiatry and Ageing Research
[10]Vassar College, Poughkeepsie, USA
[11]New York University, New York, USA
[12]University of Tübingen, Tübingen, Germany
[13]University of Surrey, Guildford, UK
[14]Aarhus University, Aarhus, Denmark
[15]Max Planck Institute for Human Development, Center for Humans & Machines, Berlin, Germany
[16]Yale University, New Haven, USA
[17]University of Bath, Bath, UK
[18]University of Oxford, Oxford, UK
[19]Baylor College of Medicine, Houston, USA
[20]Massachusetts Institute of Technology, Cambridge, USA
[*]Corresponding author: eric.schulz@tuebingen.mpg.de
[+]These authors contributed equally to this work.

ABSTRACT

Board, card, or video games have been played by virtually every individual in the world population, with both children and adults participating. Games are popular because they are intuitive and fun. These distinctive qualities of games also make them ideal as a platform for studying the mind. By being intuitive, games provide a unique vantage point for understanding the inductive biases that support behavior in more complex, ecological settings than traditional lab experiments. By being fun, games allow researchers to study new questions in cognition such as the meaning of "play" and intrinsic motivation, while also supporting more extensive and diverse data collection by attracting many more participants. We describe both the advantages and drawbacks of using games relative to standard lab-based experiments and lay out a set of recommendations on how to gain the most from using games to study cognition. We hope this article will lead to a wider use of games as experimental paradigms, elevating the ecological validity, scale, and robustness of research on the mind.

Introduction

Progress in psychological and cognitive science has been driven by the development of carefully controllable, simple, experimental paradigms that have been reused across many studies. While this approach permits precise statistical and computational modeling, it also restricts the set of answerable questions. Games present a complementary route to expand the repertoire of classic psychological tasks (1) to verify that psychological theories that have been developed in simple paradigms can explain people's behavior in more ecological settings, and (2) to ask and answer new questions about the mind, such as the form of inductive biases that support complex action, or what cognitive mechanisms support the intrinsic motivation which compels

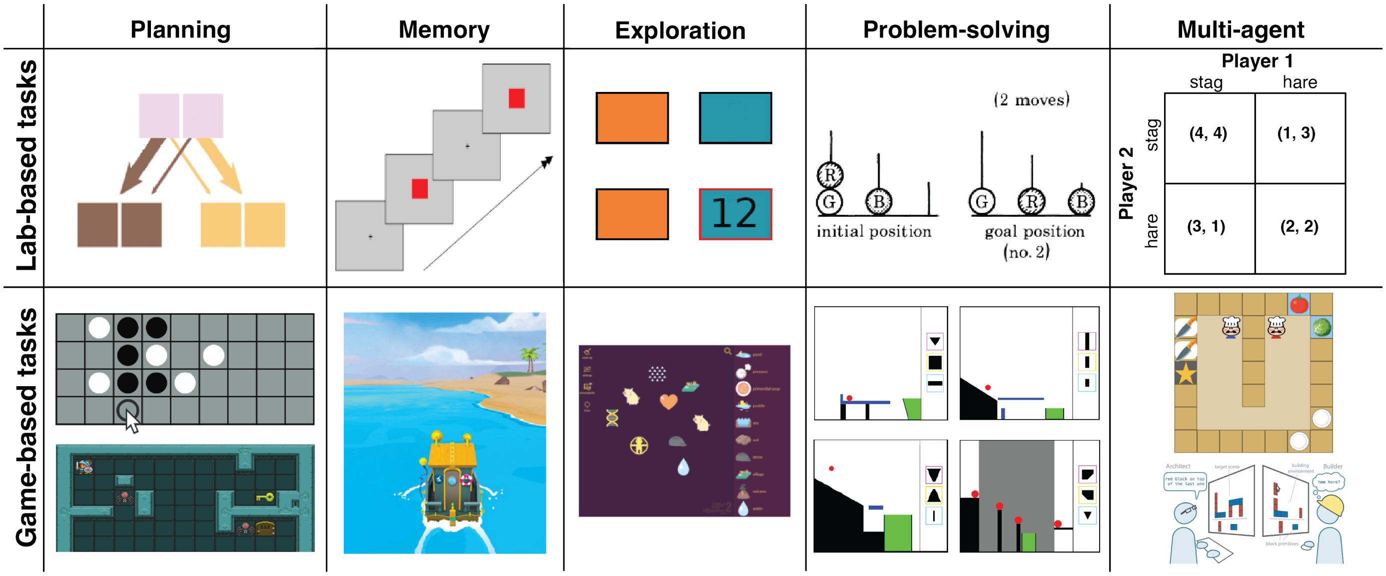

Figure 1. A comparison between classic lab-based tasks (**top**) and games (**bottom**) developed to study different facets of cognition. **Top**: from left to right, a two-step decision making task first introduced by[11], an *n*-back memory task[12], a multi-armed bandit task[13], the Towers of London problem-solving task[14], a matrix-form social coordination task[15]. **Bottom**: from left to right, the "4-in-a-row" game studied by[16] and an example of a programmatically generated video game[17,18], Sea Hero Quest[19], Little Alchemy[20], the Virtual Tools game[21], OverCooked[22,23] and multi-agent construction[24].

people to perform tasks (see Fig. 1).

Because games are frequently designed to challenge our abilities and capture our interests, they have long been central to the study of the mind[1,2]. Game playing is a popular recreational activity[3] among children and adults, across cultures[4,5], and since ancient times[6]. Yet our ability to study cognition using games has only recently been dramatically expanded by the twin advents of massive online games (which produce enormous amounts of data and can often easily be played on the phone) and advanced statistical modeling techniques. For this reason, studying how playing games affects human behavior has become increasingly important[7], and researchers have learned how to take advantage of the engaging nature of games (through "gamification") for applications in education[8,9] and therapy[10]. Despite this renewed interest in games, cognitive scientists have not yet fully embraced game-like tasks as a means to better understand the mind itself.

In this Perspectives article, we combine insights from researchers using games to study the mind across many domains and disciplines. We summarize the advantages and drawbacks of using games as a research platform, covering different types of research, and put forward recommendations on how to best use games in behavioral research. As more daily human experiences become virtual, now is a time of great potential for using games to ask and answer new questions about the brain and mind, verify small-scale theories with large-scale data, and build experiments that participants want to participate in.

Defining games

Many definitions for games have been developed over the past decades. To this day, new definitions are regularly proposed, as the nature of games is changing over time and the idea of what constitutes a game can vary[25,26]. Famously, Wittgenstein[27] even claimed that games are an ideal example of a concept that does not require a rigorous definition for its meaning to still be understood, as all games do not have a single thing in common. It is therefore difficult to find an ideal definition of games that covers all possible cases, and we instead use a definition that fits the purpose of this article - looking at how games are used for psychological research.

We, therefore, define games as "facilitators that structure player behavior and whose main purpose is enjoyment"[28]. Games structure player behavior by being intuitive, engineered environments: they reflect aspects of the real world more accurately, making them easy to interact with. Games also induce enjoyment: they are intrinsically motivating. Different tasks can be more or less game-like according to this definition. However, most classic psychological paradigms are both unintuitive (involving abstract, arbitrary rules and relying on explicit instructions to guide player behavior) and not enjoyable (requiring explicit monetary rewards to incentivize participation). Of course, this is a crude generalization and there are many games and psychological experiments that sit at different points along these axes, including gamified versions of existing psychological paradigms. These distinctions are less important in the context of this article – we want to focus instead on how developing and using tasks that are game-*like* — intuitive and enjoyable — opens up new research opportunities in psychological science, by being more ecologically valid than classic psychological paradigms but more controllable than the real world.

Potentials of games as a research platform

Games as intuitive, engineered environments

Games are designed to produce behavior approaching the complexity of the real world by being intuitive to players – that is, reflecting the assumptions that players bring to the game. In psychology research, these assumptions are often referred to as "inductive biases" – the set of assumptions that constrain and guide a learner to prefer one hypothesis over another in the absence of data. Such inductive biases have been a major focus of the field for decades, with research pointing to the importance of relational inductive biases to support flexible analogy construction[29], or object-oriented inductive biases to support perception[30].

However, the study of these inductive biases has been limited by the complexity of the tasks that are traditionally considered in psychological research. For example, in studies of decision-making, classic psychological experiments often focus on paradigms where participants must select between one of 2-4 different decisions to make (for example, pulling one of several lever arms, or choosing to go to one of 2-4 different locations), which might lead them to between 2-8 different states of the world (such as arriving at a new location). After some sequence of decisions (referred to as "planning depth" and usually only up to 4 decisions in sequence for classic studies), the participant will receive some amount of reward. Real-world decisions are significantly more complex than this, involving a nearly infinite set of different "states" a participant could end up in, as well as a nearly infinite set of different "decisions" that could be taken, with rewards sometimes not being received until after 10s or 100s of decisions.

To address this gap between psychological experiments and real-world decision-making, several recent studies have created games to study the inductive biases that allow people to reason about more complex state- and action-spaces. For example, to study more complex action-spaces, Allen et al.[21] created the "Virtual Tools Game", which requires selecting among $3x600x600$ actions – choosing one of three "tool" objects to interact with a scene, as well as the precise location on the game screen of where to place it. The central challenge people face in this game, how to cut down the possible actions to consider, much more closely reflects real-world decision-making than more traditional psychological experiments. Allen et al. found that people represent actions relationally (a relational inductive bias for action) to compress the space of actions to consider, and that such relational actions can be learned via limited amounts of trial-and-error experience[?]. To study more complex state-spaces, researchers have also used existing games such as the Atari video game suite (games like Space Invaders, Montezuma's Revenge, and Breakout), where each state is an image of a game screen ($256x256$ pixels, rather than one of a few different locations). Dubey et al.[31] found that the content of this game screen is critical. For example, when the game screen no longer consists of objects (and is instead represented as different patches of textures), people are no longer able to play the game. Furthermore, people critically rely on the existence of these objects in the state space to build relational theories about how those objects should behave (for example that keys enable you to open a door). This can then support more efficient planning and exploration[17,32].

Even with well-represented states and actions, planning in the real world often requires people to make many decisions in sequence before achieving their goal (usually referred to as "planning depth"). To study more realistic planning along this dimension,[16] developed a two-player game similar to "Tic-Tac-Toe" or "Go-Moku" in which players take turns placing tokens until one player connects four of their colored tokens in an unbroken line (see Figure 1). By working together with a mobile app company,[16] gathered data from over 1.2 million players online, as well as players in the lab. Their results confirmed that humans increased planning depth with increased expertise in this more complex planning domain, both online and in the lab. However, they also showed that online players started with worse search strategies than participants from the in-lab experiment. Therefore, there may be more opportunities to study how people improve their search strategies by studying the more general online game-playing population, as self-selected lab-based participants may already start closer to top performance.

Multi-player games can further reveal inductive biases not just for individuals, but how such inductive biases can even be shared with and shaped by other people. Unlike classical psychological experiments, which require detailed instructions to understand the task, games are designed to be intuitive enough to play without instruction. As a result, games provide an opportunity to more easily investigate cognitive phenomena across cultures[19], and to study social behaviors that rely on shared knowledge, such as cultural transmission, collective search, and other large-scale social phenomena. This is reflected in the general popularity of *multi-agent* games such as OverCooked[22,23] or Codenames[33], where playing successfully depends on having *shared* inductive biases with your teammates, or sometimes iterating on communication until such inductive biases are shared[24]. There is significant potential for further exploring how the intuitive nature of games can support such complex behaviors.

Games as enjoyable tasks

Designing experiments that participants want to participate in is important[34]. When completing an online experiment at home, there are many distractions. Attracting and maintaining participant attention is therefore crucial, and most experiments require both monetary incentives as well as "catch" trials to ensure that this is the case. Games provide an alternative mechanism for engaging participants by making participation itself naturally rewarding, or "fun".

Perhaps most uniquely, because games are naturally rewarding, they permit asking questions that are otherwise difficult to ask in regular laboratory experiments where participants are compensated for their time – for example, what intrinsically motivates people to explore a new system[20], or to persist in the presence of repeated failure[35]? Curiosity, exploration, fun and play are critical aspects of human cognition[36] which the field currently knows relatively little about, in part because these are difficult concepts to research when rewards are made explicit[37]. In particular, it is difficult to design experiments that can track all the potential kinds of exploration people can do in the real world. Games can provide agents with richer environments and make it easier to keep track of people's actions in engaging and rich environments[38]. Games provide a unique opportunity to learn more about what makes tasks fun, and how people freely behave in settings where there are no clear goals (or "play"[36]).

For example, Brändle et al[39] used the game "Little Alchemy", in which players have to combine elements to create new elements, to understand how people explore a system when there are no explicit goals. Running the game on a classical online platform commonly used for experimentation led to the necessity of compensating players for their game progress in order to motivate them to continue playing. By using data from the original mobile game, the authors were able to replicate their results in a truly intrinsic motivational setting, as players' continuation of the game must have been motivated exclusively by their enjoyment. Even in games that have a specific goal, player enjoyment can affect how they engage in a task. For example, in the Skill Lab game[40], the authors observed that compensating players via classical experimental platforms led to players exploiting some subtle flaws in the game mechanic to rapidly complete the task in an unintended way. However, by using the concept of citizen science, where players completed tasks out of intrinsic motivation, players behaved much more conscientiously and played in the spirit of the game.

Making tasks intrinsically rewarding can also lead to an increase in the amount of collected data[41] as well as the diversity of participants[42–45] including special populations such as people born with missing limbs or epilepsy patients[46,47]. However, these recruitment benefits depend upon the game being perceived as fun by these different groups. While that is not universally the case (for example, see[48] for differences in game preferences by different groups), as of 2015, regardless of gender or race, people were equally likely to play games, with about 49% of adults playing games occasionally[49] within the United States. Some games, such as chess, maintain their broad appeal across cultures to novices and experts alike and now provide enormous and exceptionally rich, diverse data sets of game-play[50–52].

This allows researchers to test theories over many more data points and participant characteristics than was previously possible. For example, the game "Sea Hero Quest" (Figure 1) collected virtual-navigation data from 4 million participants in 195 countries, giving insight into why some nations have better navigators[19], how the environment shapes future spatial skill[53] and personalized diagnostics for individuals at genetic risk of Alzheimer's disease[54].

Unlike traditional behavioral experiments, games are unique in that they exist over long time scales. Additionally, because games are enjoyable, participants often *revisit* games – sometimes over multiple days, or even multiple years – and show multiple intermediate milestones of change[55,56].

This property allows researchers to use games to study the acquisition of expertise and related representational and strategy changes, which can take years to develop[57]. For example, experienced chess players perceive and remember mid-game positions as larger chunks than novices[58,59]. However, such studies usually rely on examining differences between individuals who already have different amounts of expertise – novices and experts, for example. To study how expertise is acquired, people need to be observed over many more sessions than is practical for in-lab testing. To overcome this limitation,[60] worked with game designers "Preloaded" to develop the game "Axon" which requires players to guide a neuron from connection to connection, by rapidly clicking on potential targets. By measuring the performance of individual players from the very first time they interacted with the game, the authors found that while practice improved all players' performance, it did not affect all players equally. Some players were better from their very first attempt, and their practice produced steeper learning curves than for players who had initially low scores. This would not be possible to study if only looking at novices and experts – the whole time course is needed to understand the influence of practice.

Potential pitfalls of games as a research platform

Despite the enormous potential of games in psychological research, they also come with important drawbacks relative to classical psychological paradigms. These can be broadly categorized into two topics: unique challenges of experimental design with games, and data collection & analysis.

Experimental design Experimental design with games is challenging because experimenters have less control over how a game is presented to participants. This manifests in two major ways.

First, games are engineered to be fun and to attract player attention. This sometimes requires superfluous "bells and whistles" which may or may not affect the cognitive process under investigation. Since researchers cannot always precisely control the presentation of a game to a participant, it is more difficult to mitigate potential experimental confounds introduced by these bells and whistles. This may dampen the generalizability of results obtained using existing games, as the specific behaviors or patterns observed in a game may be a result of the incentive structure of that specific game, rather than reflecting general principles of cognition.

Second, games exist outside of the confines of an experiment. This means that games may also suffer from greater variability between subjects' prior experiences of the task, which may also affect their performance. These differences can also quickly accumulate as different individuals may progress through the game at different rates – changing the nature and quantity of their interactions. Between sessions, some players may even participate in activities that highly overlap with the game, such as other games with related skill requirements, or even discuss the game in online communities. This may result in non-representative data, which could lead to further problems with data analysis.

Some of these limitations may be insurmountable when using an existing popular game for an experiment. However, we propose two mitigation strategies: validating game-based results using a carefully controlled psychological experiment or other complementary games[61,62], or creating your own game where direct experimental control is possible[19,21,63]. We detail these considerations further in "How to use games as a research platform".

Data collection & analysis Once a game-based experiment has been designed, collecting and analyzing data also presents unique challenges relative to classic psychological tasks, which we detail below.

First, data collection with games can be challenging. Game designers may not always be willing to share data. However, setting up the infrastructure needed to both collect data and track player progress across time can be daunting for academics, as these are not features supported by most standard online experimental platforms. Thankfully, special infrastructures – like virtual labs – have been developed, so that researchers have all the usability and functionality they need, while subjects can still easily access and participate in the experiments at their own pace[40,64–66].

Once the data has been obtained, it may also require different analyses relative to data obtained with classic psychology experiments. In particular, given the complexity of game-based data, there is a risk of using ad-hoc measures to make conclusions that are not driven by theoretical considerations. This risk may be exacerbated given that many games were not developed to measure individual cognitive processes in the first place. Game-based data sets can also be very large, increasing the risk of finding "statistically significant" results from such ad-hoc measures, which may not generalize. However, this does not mean that newly derived measures for analyzing data from games cannot be validated. If such derived measures can also predict everyday behavior (i.e., they have *predictive validity*), or they produce similar conclusions to established measurements (i.e., they have *concurrent validity*), researchers can be more confident in their conclusions. These considerations for validating game-based measurements may also constrain the search for good candidate measures and therefore reduce their arbitrariness overall.

Even when the measures are known and well-validated, games are often more complex than traditional psychology experiments (by better resembling real-world behavior), and may therefore be difficult to model statistically or computationally[16]. Some researchers have turned to contemporary techniques from machine learning and AI to aid with modeling complex game stimuli (such as how to computationally represent pixels on a screen[31,67]). Since more complex models, such as multi-layer neural networks or planning algorithms from AI, can be set up in various different ways, comparing models in games may require a shift from simply comparing one model to another towards comparing classes of different algorithms[68]. For example, if all models that describe human planning well require a particular set of features to calculate the value of states, then it is likely that these features matter for human planning independent of the particular model class. For example,[16] showed that (amongst others) tree search, tantamount to mentally simulating consequences of available actions, and feature dropping, akin to spatial- and feature-based attention, are necessary model features to account for choices in a sequential decision making task.

How to use games as a research platform

What if you would like to use a game in your own research? Here, we provide a guide for how to maximally benefit from using games as a research platform.

Independent of their origin, using games requires making several crucial decisions, such as how sparse players' rewards should be, how complex the game should be (e.g., by focusing on complexity in different domains like strategy or input), and how one should set up the game's progression (or curriculum) from level to level. Similarly, since games provide a unique opportunity to study inductive biases, it is essential to think about which priors people could have for a given game. Finally, it is important to think about whether you want to gamify individual cognitive constructs such as exploration or planning abilities

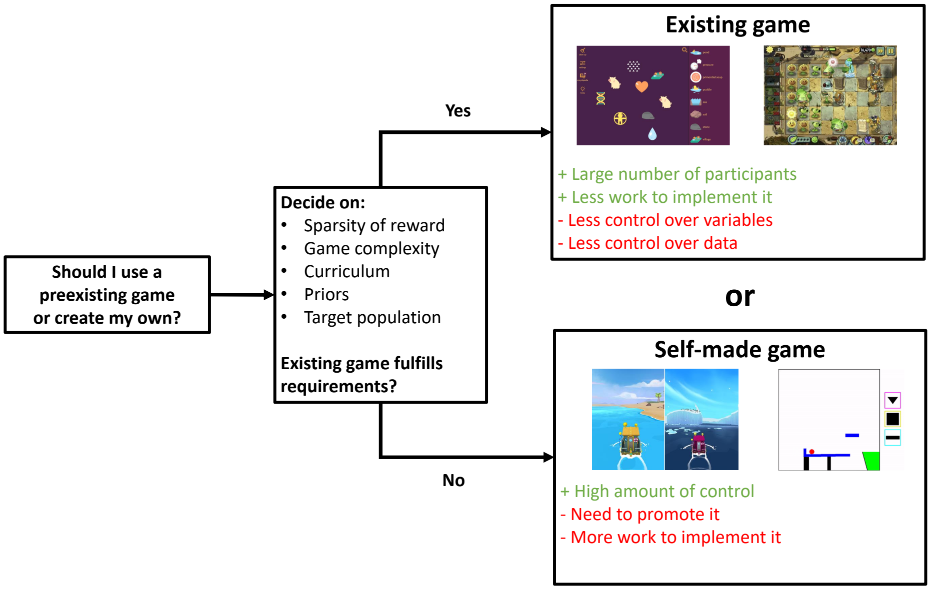

Figure 2. Decision criteria, advantages and drawbacks of using existing games and self-made games. **Top**: from left to right: Little Alchemy 2[20], Use your Brainz[69]. **Bottom**: from left to right: Sea Hero Quest[19], the Virtual Tools game[21]

or if you want to pursue a more portfolio-based approach of using games to study multiple psychological constructs[40,64]. We believe that these choices and considerations should be made with clear hypotheses in mind about the underlying cognitive mechanisms that researchers want to assess[70].

After deciding which hypotheses to test with a game, you have to decide where your game is going to come from. Broadly, there are two options: you can create and host your own game, or you can use a pre-existing game (like Chess, Angry Birds, etc. (see Figure 2). While creating your own game gives you more control over specific game parameters, and is not hard to do with current game development tools such as PyMunk[71] or Unity[72], it requires you to think about where to publish the game to recruit enough participants. Using an existing popular game has the advantage of having an already existing player base, and therefore a significant amount of data to analyze. However, even if available, these data may not be well-suited to answer a particular research question. It can also be challenging to convince a game company to provide access to their data. In our experience, smaller companies or even individual developers are more likely to be interested in collaboration. To increase your chances of a positive response, we recommend introducing yourself to your institution and area of research, expressing enthusiasm for the game, and stating in plain language the research question you hope to answer using the game's data. We also recommend emphasizing the benefits of collaboration to the game creator (e.g. that studying the game might lead to more publicity for the game itself).

Between these two extremes, researchers can also partner with game developers to make custom games or to gamify existing classical experimental paradigms[73,74]. Thereby, researchers keep more control over the game than by using pre-existing data, while profiting from the expertise and distribution workflows of professional game developers. To support partnership with game developers, several conferences exist which bring researchers and programmers together. [1]

If using a pre-existing game or partnering with a game developer, maintaining positive relationships with game creators is essential. However, there can be conflicting goals especially when decisions by game creators affect how stimuli are presented and data are collected for research purposes. In our experience, the best way to overcome these challenges is to either (a) work with smaller companies or individual developers who may be more amenable to constructive solution-finding with researchers, (b) work with established games that are not undergoing major development, or (c) obtain permission to make a copy of the game for research purposes, and hosting this copy separately to collect data through research platforms such as Prolific or Amazon Mechanical Turk.

To attract a target population with your game, at least two points have to be considered. First, the appearance of games should be adapted to the target population. For example, a game designed for children should look very different from a

[1]BrainPlay (CMU; https://www.cmu.edu/ni/events/brain-play-2022.html), neurodiversity in Tech (UCSD; https://pong-center.ucsd.edu/internship/), and ReGame-XR Summer Internship (Northeastern; https://regamexr.sites.northeastern.edu/internship/)

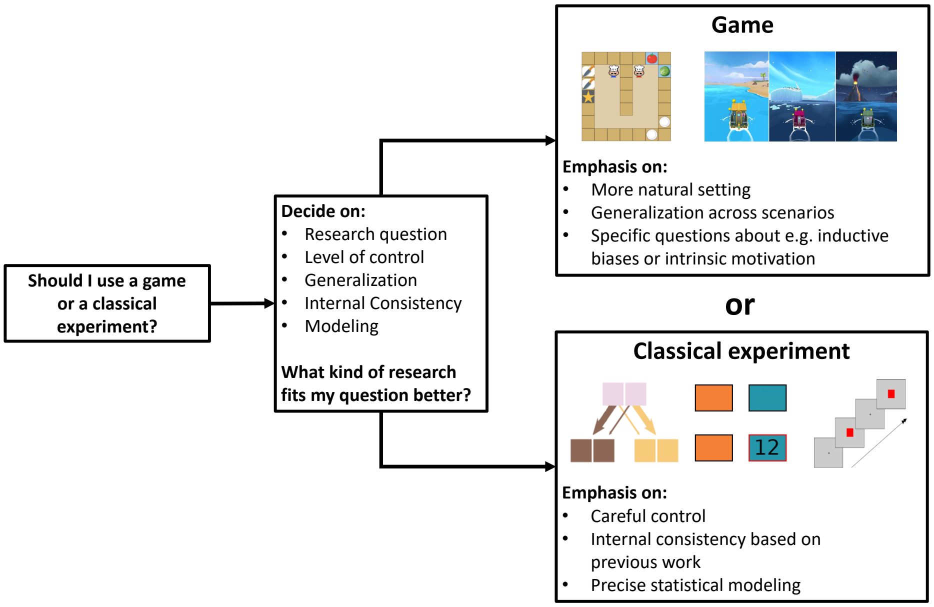

Figure 3. Decision criteria to choose between games and classical experiments. **Top**: from left to right: OverCooked[22,23], Sea Hero Quest[19]. **Bottom**: from left to right: a two-step decision-making task first introduced by[11], a multi-armed bandit task[13], an *n*-back memory task[12]

game that is designed for strategy game enthusiasts. Second, there can be a selection bias in a given participant pool because different games might attract players with different characteristics (for example, see[48]). If a diverse participant pool is important, standard recruitment platforms such as Mechanical Turk and Prolific are a viable option because they allow researchers to control these characteristics. However, if intrinsic motivation is important, standard platforms can be problematic and we suggest recruiting participants over social media, putting the game on different app stores, or –if necessary– working with a game company directly. Publicly announcing that results are going to be used for research purposes may also call citizen scientists into action[75], thus further engaging and diversifying participants[62]. Moreover, researchers can also partner with organizations and charities to reach out to different communities to play games.

Finally, we give some advice on how to analyze data collected in games. We suggest storing the ultimately larger data sets in a database (for example, SQL or Mongo), which gives control over the data's organization. Because of the size of these datasets, it is particularly important to derive predictions a priori and focus on the variables relevant for testing them to avoid getting lost in endless analyses or finding spurious statistical effects. We also recommend reporting effect sizes in addition to measures of significance, which can be misleading in large data sets, to gauge the relative importance of an effect. Moreover, since different players have different exposures to a game, it may be necessary to condition analyses on the number of trials or levels a particular player has played. Alternatively, one could implement a "test level" at the beginning of each session, to check for a change in skill level.

Analyzing data from games is usually more difficult than analyzing data from experiments because defining an unbiased likelihood function becomes challenging and because gradient descent methods struggle. We suggest sampling-based methods for log-likelihood estimation (e.g., inverse binomial sampling[76]) and global parameter optimization techniques (e.g., Bayesian optimization[77]), respectively, to circumvent these difficulties.

Besides these theoretical risks, legal risks (e.g., data protection) and financial risks (e.g., funding of the study) can be more pronounced as compared to studies conducted in the lab, for which default procedures may already be available. We, therefore, recommend thorough planning of studies and talking to your legal and finance department in case of any ambiguities.

In this article, we have tried to distinguish games from classical experiments (see Fig. 3). However, the two approaches are not mutually exclusive and we think that research can profit from combining the two. Several of the characteristics differentiating games from experiments are related to the tension between internal and external validity inherent to any empirical study[78]. For example, whereas games allow researchers to test theories in more natural settings, they are less controllable. The reverse is true for experiments. In order to make statements that are internally consistent, but also generalize well across scenarios, we suggest combining games and experiments. Like previous research[40,79,80], we suggest two strategies for merging game-based and experiment-based research. In the bottom-up strategy, a researcher starts by demonstrating a cognitive mechanism in an experiment and then tries to generalize that mechanism to a more complex game – potentially considering boundary conditions. In the top-down strategy, the researcher can start by demonstrating a mechanism in a game and then validate it using carefully controlled and simplified experiments. These more traditional experiments can even be added to the end of online games, where previous work has found that players can still be eager to contribute to these perhaps less exciting tasks after having played a game[40]. We do not believe that games should replace experiments but rather that games and in-lab experiments will work best in tandem.

Future questions that can be addressed by games

We have argued for the usefulness of games to study many cognitive domains. Yet many questions remain that we believe could be addressed using games in future research including intrinsic motivation and the meaning of “fun", cognitive processes, neuronal processes, game-playing in animals, and not just how people *play* games, but rather how people *create* games.

People’s willingness to spend hours at a game is deeply tied to their intrinsic motivation to play. Why do we enjoy playing games at all? What compels someone to play a game for the first time, and then to continue playing it to reach deep expertise? Why do we pursue goals despite no obvious extrinsic reward? Games are uniquely positioned to help us answer these questions about our innermost motivations.

We believe that games can advance psychological research. However, a thorough study of whether the superfluous “bells and whistles” that games make use of to attract players change cognitive processes could advance and improve the usage of games as research paradigms. Especially focusing on which parts of game design influence different aspects of players’ behavior apart from engagement would advance future study designs.

Games can not only advance psychological research of human behavior but can also be integrated into neuroscientific research. For example,[81] found that neural responses to success and failure in games were modulated by active vs. vicarious game-playing. More generally, by letting people play games during the simultaneous acquisition of neural recordings, it is possible to study neuronal processes in more naturalistic environments[47,82].

Beyond human game-play, we also believe that studying how animals play games, and what “fun" might mean to other species, will be a fruitful area for future research. While there have been some prior studies of game-play in primates[83,84], there is much room for future research into what kinds of activities would be considered games for animals.

Finally, we believe that games are also interesting for how they are *developed*. Game design is a highly creative endeavor, requiring a sophisticated understanding of the beliefs and knowledge of the game players. How do people create games from those as simple as tic-tac-toe to those as complex as Minecraft? We believe there are many open questions here to explore.

Conclusion

We have argued that truly understanding the mind requires a paradigm shift away from only using highly controlled and simplified experiments and towards the rich landscape of studying learning, decision-making, language understanding, and comparative cognition afforded by games as a research platform. Research on the mind can benefit greatly from the additional insights and improved understanding that can come from this shift. At the same time, these benefits may falter without minimizing the potential pitfalls associated with the increased complexity found in many games. We believe that these can be mitigated by making sure that games are designed to test particular hypotheses about human behavior and by making sure that computational modeling is sufficiently tailored to the games in question. Ultimately, we believe that reverse engineering normally works best when people are put in environments to which they are adapted (i.e., “engineered”), and well-designed games can offer such environments.

Acknowledgements

We thank Abhilasha Ashok Kumar and Yann Harel for helpful discussions.

References

1. Newell, A., Simon, H. A. *et al. Human problem solving*, vol. 104 (Prentice-hall Englewood Cliffs, NJ, 1972).
2. Gobet, F., de Voogt, A. & Retschitzki, J. *Moves in Mind: The Psychology of Board Games* (2004).
3. Gough, C. Number of gamers worldwide 2021. *Statista* **13**, 2020 (2019).
4. Suchow, J. W., Griffiths, T. & Hartshorne, J. K. Workshop on scaling cognitive science. In *CogSci* (2020).
5. Brändle, F., Allen, K. R., Tenenbaum, J. & Schulz, E. Using games to understand intelligence. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 43 (2021).

6. DePaulis, T. Board games before ur? *Board Game Stud. J.* **14**, 127–144 (2020).
7. Greitemeyer, T. & Mügge, D. O. Video Games Do Affect Social Outcomes: A Meta-Analytic Review of the Effects of Violent and Prosocial Video Game Play. *Pers. Soc. Psychol. Bull.* **40**, 578–589, DOI: 10.1177/0146167213520459 (2014).
8. Bertram, L. Digital learning games for mathematics and computer science education: The need for preregistered rcts, standardized methodology, and advanced technology. *Front. Psychol.* 2127, DOI: 10.3389/fpsyg.2020.02127 (2020).
9. Manzano-León, A. *et al.* Between level up and game over: A systematic literature review of gamification in education. *Sustainability* **13**, DOI: 10.3390/su13042247 (2021).
10. Fleming, T. M. *et al.* Serious games and gamification for mental health: Current status and promising directions. *Front. Psychiatry* **7**, DOI: 10.3389/fpsyt.2016.00215 (2017).
11. 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).
12. Kirchner, W. K. Age differences in short-term retention of rapidly changing information. *J. experimental psychology* **55**, 352 (1958).
13. Gershman, S. J. Uncertainty and exploration. *Decision* **6**, 277 (2019).
14. Shallice, T. Specific impairments of planning. *Philos. Transactions Royal Soc. London. B, Biol. Sci.* **298**, 199–209 (1982).
15. Costa-Gomes, M., Crawford, V. P. & Broseta, B. Cognition and behavior in normal-form games: An experimental study. *Econometrica* **69**, 1193–1235 (2001).
16. van Opheusden, B. *et al.* Expertise increases planning depth in human gameplay. *Nature* 1–6 (2023).
17. Tsividis, P. A. *et al.* Human-level reinforcement learning through theory-based modeling, exploration, and planning. *arXiv preprint arXiv:2107.12544* (2021).
18. Schaul, T. A video game description language for model-based or interactive learning. In *2013 IEEE Conference on Computational Inteligence in Games (CIG)*, 1–8 (IEEE, 2013).
19. Coutrot, A. *et al.* Global determinants of navigation ability. *Curr. Biol.* **28**, 2861–2866 (2018).
20. Brändle, F., Binz, M. & Schulz, E. Exploration beyond bandits. *PsyArXiv* (2021).
21. Allen, K. R., Smith, K. A. & Tenenbaum, J. B. Rapid trial-and-error learning with simulation supports flexible tool use and physical reasoning. *Proc. Natl. Acad. Sci.* **117**, 29302–29310 (2020).
22. Wang, R. E. *et al.* Too many cooks: Coordinating multi-agent collaboration through inverse planning. *CoRR* **abs/2003.11778** (2020).
23. Carroll, M. *et al.* On the utility of learning about humans for human-ai coordination. *Adv. neural information processing systems* **32** (2019).
24. McCarthy, W. P., Hawkins, R. D., Wang, H., Holdaway, C. & Fan, J. E. Learning to communicate about shared procedural abstractions. *CoRR* **abs/2107.00077** (2021).
25. Arjoranta, J. How to define games and why we need to. *The Comput. Games J.* **8**, 109–120 (2019).
26. Stenros, J. The game definition game: A review. *Games Cult.* **12**, 499–520 (2017).
27. Wittgenstein, L. *Philosophical Investigations: The German text, with an English translation by G.E.M. Anscombe, P.M.S. Hacker and Joachim Schulte* (Blackwell Publishing Ltd, 2009), 4 edn.
28. Aarseth, E. I fought the law: Transgressive play and the implied player. In *DiGRA &#3907 - Proceedings of the 2007 DiGRA International Conference: Situated Play* (The University of Tokyo, 2007).
29. Gentner, D. Structure-mapping: A theoretical framework for analogy. *Cogn. science* **7**, 155–170 (1983).
30. Spelke, E. S., Breinlinger, K., Macomber, J. & Jacobson, K. Origins of knowledge. *Psychol. review* **99**, 605 (1992).
31. Dubey, R., Agrawal, P., Pathak, D., Griffiths, T. L. & Efros, A. A. Investigating human priors for playing video games. *CoRR* **abs/1802.10217** (2018).
32. Pouncy, T. & Gershman, S. J. Inductive biases in theory-based reinforcement learning. *Cogn. Psychol.* (2022).
33. Kumar, A. A., Steyvers, M. & Balota, D. A. Semantic memory search and retrieval in a novel cooperative word game: A comparison of associative and distributional semantic models. *Cogn. Sci.* **45**, e13053 (2021).
34. Jun, E., Hsieh, G. & Reinecke, K. Types of motivation affect study selection, attention, and dropouts in online experiments. *Proc. ACM on Human-Computer Interact.* **1**, 1–15 (2017).

35. Leonard, J. A., Kleiman-Weiner, M., Lee, Y., Tenenbaum, J. & Schulz, L. Preschoolers and infants calibrate persistence from adult models. In *CogSci* (2017).

36. Chu, J. & Schulz, L. E. Play, curiosity, and cognition. *Annu. Rev. Dev. Psychol.* **2**, 317–343 (2020).

37. Murayama, K. A reward-learning framework of knowledge acquisition: An integrated account of curiosity, interest, and intrinsic-extrinsic rewards. *Psychol. Rev.* **129**, 175–198, DOI: 10.1037/rev0000349 (2022).

38. Kosoy, E. *et al.* Exploring exploration: Comparing children with rl agents in unified environments. *arXiv preprint arXiv:2005.02880* (2020).

39. Brändle, F., Stocks, L. J., Tenenbaum, J., Gershman, S. J. & Schulz, E. Intrinsic exploration as empowerment in a richly structured online game, DOI: 10.31234/osf.io/ybs7g (2022).

40. Pedersen, M. K. *et al.* Cognitive abilities in the wild: Population-scale game-based cognitive assessment. *arXiv preprint arXiv:2009.05274* (2020).

41. Hartshorne, J. K., de Leeuw, J. R., Goodman, N. D., Jennings, M. & O'Donnell, T. J. A thousand studies for the price of one: Accelerating psychological science with pushkin. *Behav. research methods* **51**, 1782–1803 (2019).

42. Thorne, H. T., Smith, J. J., Morgan, P. J., Babic, M. J. & Lubans, D. R. Video game genre preference, physical activity and screen-time in adolescent boys from low-income communities. *J. adolescence* **37**, 1345–1352 (2014).

43. Jansz, J., Avis, C. & Vosmeer, M. Playing the sims2: An exploration of gender differences in players' motivations and patterns of play. *New Media & Soc.* **12**, 235–251 (2010).

44. Hartshorne, J. K., Tenenbaum, J. B. & Pinker, S. A critical period for second language acquisition: Evidence from 2/3 million english speakers. *Cognition* **177**, 263–277 (2018).

45. Awad, E. *et al.* The moral machine experiment. *Nature* **563**, 59–64, DOI: 10.1038/s41586-018-0637-6 (2018).

46. Allen, K. R. *et al.* Lifelong learning of cognitive strategies in physical problem-solving: the effect of embodied experience. *bioRxiv* (2021).

47. Ashmaig, O. *et al.* A platform for cognitive monitoring of neurosurgical patients during hospitalization. *Front. Hum. Neurosci.* **15**, 702, DOI: 10.3389/fnhum.2021.726998 (2021).

48. Andrews, G. Gameplay, gender, and socioeconomic status in two american high schools. *E-learning Digit. Media* **5**, 199–213 (2008).

49. Pew Research Center. Gaming and gamers. Tech. Rep., Pew Research Center, Washington, D.C. (2015).

50. Chassy, P. & Gobet, F. Risk taking in adversarial situations: Civilization differences in chess experts. *Cognition* **141**, 36–40 (2015).

51. Morsch, F. & Feist, M. Fritz (version 12). *Hamburg, Ger. Chessbase* (2009).

52. Holdaway, C. & Vul, E. Risk-taking in adversarial games: What can 1 billion online chess games tell us? In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 43 (2021).

53. Coutrot, A. *et al.* Entropy of city street networks linked to future spatial navigation ability. *Nature* **604**, 104–110 (2022).

54. Coughlan, G. *et al.* Toward personalized cognitive diagnostics of at-genetic-risk alzheimer's disease. *Proc. Natl. Acad. Sci.* **116**, 9285–9292 (2019).

55. Chu, Y. & MacGregor, J. N. Human performance on insight problem solving: A review. *The J. Probl. Solving* **3**, 6 (2011).

56. Gick, M. L. & Holyoak, K. J. Analogical problem solving. *Cogn. psychology* **12**, 306–355 (1980).

57. de Groot, A. D. *Het denken van den schaker: een experimenteel-psychologische studie* (, 1946).

58. Chase, W. G. & Simon, H. A. Perception in chess. *Cogn. Psychol.* **4**, 55–81, DOI: 10.1016/0010-0285(73)90004-2 (1973). Number: 1.

59. Gobet, F. *et al.* Chunking mechanisms in human learning. *Trends Cogn. Sci.* **5**, 236–243 (2001).

60. Stafford, T. & Dewar, M. Tracing the trajectory of skill learning with a very large sample of online game players. *Psychol. Sci.* **25**, 511–518, DOI: 10.1177/0956797613511466 (2014). PMID: 24379154, https://doi.org/10.1177/0956797613511466.

61. Eisenberg, I. W. *et al.* Uncovering the structure of self-regulation through data-driven ontology discovery. *Nat. communications* **10**, 1–13 (2019).

62. Rafner, J. *et al.* Digital games for creativity assessment: Strengths, weaknesses and opportunities. *Creat. Res. J.* **34**, 28–54 (2022).

63. Schulz, E., Bertram, L., Hofer, M. & Nelson, J. D. Exploring the space of human exploration using entropy mastermind (2019).

64. Brown, H. R. *et al.* Crowdsourcing for cognitive science–the utility of smartphones. *PloS one* **9**, e100662 (2014).

65. Griffiths, T. L. Manifesto for a new (computational) cognitive revolution. *Cognition* **135**, 21–23 (2015).

66. Almaatouq, A. *et al.* Scaling up experimental social, behavioral, and economic science. *Comput. Soc. Sci. Lab at Univ. Pennsylvania* 1–8 (2021).

67. Cross, L., Cockburn, J., Yue, Y. & O'Doherty, J. P. Using deep reinforcement learning to reveal how the brain encodes abstract state-space representations in high-dimensional environments. *Neuron* **109**, 724–738 (2021).

68. Peterson, J. C., Bourgin, D. D., Agrawal, M., Reichman, D. & Griffiths, T. L. Using large-scale experiments and machine learning to discover theories of human decision-making. *Science* **372**, 1209–1214 (2021).

69. Shute, V. J., Wang, L., Greiff, S., Zhao, W. & Moore, G. Measuring problem solving skills via stealth assessment in an engaging video game. *Comput. Hum. Behav.* **63**, 106–117 (2016).

70. Rafferty, A. N., Zaharia, M. & Griffiths, T. L. Optimally designing games for behavioural research. *Proc. Royal Soc. A: Math. Phys. Eng. Sci.* **470**, 20130828 (2014).

71. Pymunk. https://pymunk.org.

72. Juliani, A. *et al.* Unity: A general platform for intelligent agents (2020). 1809.02627.

73. Mazarakis, A. & Bräuer, P. Gamification is working, but which one exactly? results from an experiment with four game design elements. *Int. J. Human-Computer Interact.* **39**, 612–627, DOI: 10.1080/10447318.2022.2041909 (2023).

74. Sage & Hodges, N. How to gamify cognitive tasks to increase participant engagement.

75. Silvertown, J. A new dawn for citizen science. *Trends Ecol. & Evol.* **24**, 467–471, DOI: 10.1016/j.tree.2009.03.017 (2009).

76. van Opheusden, B., Acerbi, L. & Ma, W. J. Unbiased and efficient log-likelihood estimation with inverse binomial sampling. *PLoS computational biology* **16**, e1008483 (2020).

77. Acerbi, L. & Ma, W. J. Practical bayesian optimization for model fitting with bayesian adaptive direct search. *Adv. neural information processing systems* **30** (2017).

78. Gravetter, F. J. & Forzano, L.-A. B. *Research methods for the behavioral sciences* (Wadsworth, Australia ; Belmont, CA, 2012), 4th ed edn.

79. Lewkowicz, D. J. The Concept of Ecological Validity: What Are Its Limitations and Is It Bad to Be Invalid? *Infancy* **2**, 437–450, DOI: 10.1207/S15327078IN0204_03 (2001). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/S15327078IN0204_03.

80. Lickliter, R. & Bahrick, L. E. The Salience of Multimodal Sensory Stimulation in Early Development: Implications for the Issue of Ecological Validity. *Infancy* **2**, 451–467, DOI: 10.1207/S15327078IN0204_04 (2001). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/S15327078IN0204_04.

81. Kätsyri, J., Hari, R., Ravaja, N. & Nummenmaa, L. Just watching the game ain't enough: striatal fmri reward responses to successes and failures in a video game during active and vicarious playing. *Front. Hum. Neurosci.* **7**, DOI: 10.3389/fnhum.2013.00278 (2013).

82. Mathiak, K. & Weber, R. Toward brain correlates of natural behavior: fmri during violent video games. *Hum. Brain Mapp.* **27**, 948–956, DOI: https://doi.org/10.1002/hbm.20234 (2006).

83. Yang, Q. *et al.* Monkey plays pac-man with compositional strategies and hierarchical decision-making. *eLife* **11**, e74500, DOI: 10.7554/eLife.74500 (2022).

84. Haroush, K. & Williams, Z. M. Neuronal prediction of opponent's behavior during cooperative social interchange in primates. *Cell* **160**, 1233–1245, DOI: 10.1016/j.cell.2015.01.045 (2015).