nature human behaviour

Perspective

https://doi.org/10.1038/s41562-024-01878-9

Using games to understand the mind

Received: 15 June 2022

Accepted: 3 April 2024

Published online: 21 June 2024

Check for updates

Kelsey Allen[1,21], Franziska Brändle[2,21], 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[8,9,10], Mark K. Ho[6], Joshua R. de Leeuw[11], Wei Ji Ma[12], Kou Murayama[10], 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] & Eric Schulz[2] ✉

Board, card or video games have been played by virtually every individual in the world. Games are popular because they are intuitive and fun. These distinctive qualities of games also make them ideal for studying the mind. By being intuitive, games provide a unique vantage point for understanding the inductive biases that support behaviour in more complex, ecological settings than traditional laboratory 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 the advantages and drawbacks of using games relative to standard laboratory-based experiments and lay out a set of recommendations on how to gain the most from using games to study cognition. We hope this Perspective will lead to a wider use of games as experimental paradigms, elevating the ecological validity, scale and robustness of research on the mind.

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. Although this approach permits precise statistical and computational modelling, it also restricts the set of answerable questions. Games present a complementary route to expand the repertoire of classic psychological tasks, allowing researchers to (1) verify that psychological theories that have been developed in simple paradigms can explain people's behaviour in more ecological settings, and (2) 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 that compels people to perform tasks (Fig. 1).

Because games are frequently designed to challenge our abilities and capture our interests, simple games were traditionally used to study 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]. Our ability to study cognition using games has 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 a phone) and advanced statistical modelling techniques. For this reason,

[1]DeepMind, London, UK. [2]Max Planck Institute for Biological Cybernetics, Tübingen, Germany. [3]Stanford University, Stanford, CA, USA. [4]Harvard University, Cambridge, MA, USA. [5]University of California, Berkeley, Berkeley, CA, USA. [6]Princeton University, Princeton, NJ, USA. [7]Boston College, Boston, MA, USA. [8]University College London, London, UK. [9]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, London, UK. [10]University of Tübingen, Tübingen, Germany. [11]Vassar College, Poughkeepsie, NY, USA. [12]New York University, New York, NY, USA. [13]University of Surrey, Guildford, UK. [14]Aarhus University, Aarhus, Denmark. [15]Center for Humans and Machines, Max Planck Institute for Human Development, Berlin, Germany. [16]Yale University, New Haven, CT, USA. [17]University of Bath, Bath, UK. [18]University of Oxford, Oxford, UK. [19]Baylor College of Medicine, Houston, TX, USA. [20]Massachusetts Institute of Technology, Cambridge, MA, USA. [21]These authors contributed equally: Kelsey Allen, Franziska Brändle. ✉e-mail: eric.schulz@tuebingen.mpg.de

 https://doi.org/10.1038/s41562-024-01878-9

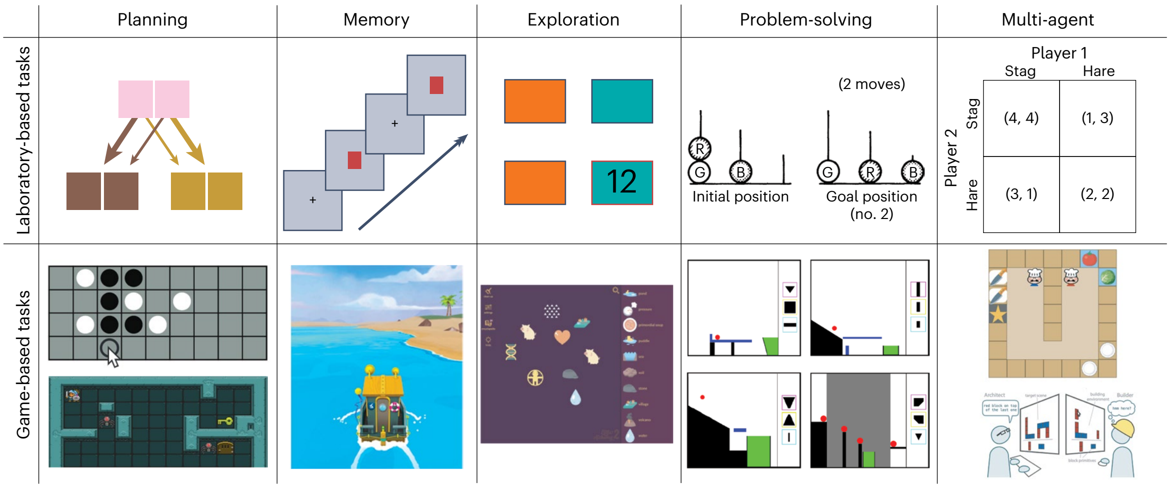

Fig. 1 | A comparison between classic laboratory-based tasks and games developed to study different facets of cognition. Top, from left to right, a two-step decision-making task first introduced by Daw et al.[77], an *n*-back memory task[78], a multi-armed bandit task[79], the Towers of London problem-solving task[80] and a matrix-form social coordination task[81]. Bottom, from left to right, the '4-in-a-row' game studied by van Opheusden et al.[21] and an example of a programmatically generated video game[20,82], Sea Hero Quest[22], *Little Alchemy 2* (ref. 33), the Virtual Tools game[17], a cooking game inspired by *OverCooked*[23,24] and a multi-agent construction game[26]. For laboratory tasks (problem-solving), image adapted with permission of The Royal Society, from ref. 80; permission conveyed through Copyright Clearance Center, Inc. For game-based tasks, images adapted from ref. 21, Springer Nature Ltd (planning, top); and adapted with permission from ref. 20, P. A. Tsividis et al. (planning, bottom); ref. 22, Glitchers Ltd and Alzheimer's Research UK (memory); ref. 33, Recloak/Jakub Kozioł (exploration); ref. 17, PNAS (problem-solving); ref. 23, Wiley (multi-agent, top); ref. 26, W. P. McCarthy et al. (multi-agent, bottom).

studying how playing games affects human behaviour 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]. Cognitive scientists are similarly well positioned to embrace these new developments in online games to better understand the mind itself.

In this Perspective, 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 behavioural 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 people want to participate in.

Defining games in virtual environments

Many definitions of 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[11,12]. Famously, Wittgenstein[13] claimed that games are an ideal example of a concept that does not require a rigorous definition for its meaning to be understood, as games do not all 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 Perspective—looking at how virtual games are used for psychological research.

We therefore define games as "facilitators that structure player behavior and whose main purpose is enjoyment"[14]. Games structure player behaviour by being intuitive, engineered environments: they reflect aspects of the real world 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 behaviour) and unenjoyable (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 Perspective—we want to focus instead on how developing and using virtual 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 behaviour 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[15] or object-oriented inductive biases to support perception[16].

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 one of two to four different decisions (or 'actions') to make (for example, pulling one of several lever arms or choosing to go to one of two to four different locations (Fig. 1, top left)), which might lead them to between two and eight different states of the world (such as arriving at a new location). After some sequence of actions (referred to as 'planning depth' and usually only up to four actions in sequence for classic studies), the participant receives some amount of reward. Real-world decisions are substantially 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 actions that could be taken, with rewards sometimes not being received until after tens or hundreds of actions.

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.[17] created the Virtual Tools game, which requires selecting among 3 × 600 × 600 actions—choosing one of three 'tool' objects to interact with a scene, as well as the precise location on the game screen in which 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[17]. To study more complex state spaces, researchers have also used existing games such as the Atari video game suite (games such as *Space Invaders*, *Montezuma's Revenge* and *Breakout*), where each state is an image of a game screen (256 × 256 pixels, rather than one of a few different locations). Dubey et al.[18] 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[19,20].

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, van Opheusden et al.[21] developed a two-player game similar to noughts and crosses or *go-moku*, in which players take turns placing tokens until one player connects four of their coloured tokens in an unbroken line (Fig. 1). By working together with a mobile app company, van Opheusden et al.[21] gathered data from over 1.2 million players online, as well as players in the laboratory. Their results confirmed that humans increased planning depth with increased expertise in this more complex planning domain, both online and in the laboratory. However, they also showed that online players started with worse search strategies than participants from the in-laboratory 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 laboratory-based participants may already start closer to top performance.

Multi-player games can further reveal not just inductive biases for individuals but also how such inductive biases can 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[22] and to study social behaviours 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*[23,24] or *Codenames*[25], in which playing successfully depends on having shared inductive biases with your teammates or sometimes iterating on communication until such inductive biases are shared[26]. There is substantial potential for further exploring how the intuitive nature of games can support such complex behaviours.

Games as enjoyable tasks

Designing experiments that participants want to participate in is important[27]. When completing an online experiment at home, participants have many distractions. Attracting and maintaining participant attention is therefore crucial, and most experiments require both monetary incentives and 'catch' trials to ensure that this is achieved. 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[28] or to persist in the presence of repeated failure[29]? Curiosity, exploration, fun and play are critical aspects of human cognition[30] that the field currently knows relatively little about, in part because these are difficult concepts to research when rewards are made explicit[31]. 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 make it easier to keep track of people's actions in engaging and rich environments[32]. 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 (that is, play[30]).

For example, Brändle et al.[33] used the game *Little Alchemy 2*, 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 investigate players' decisions 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[34], 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, when the authors used 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[35] as well as the diversity of participants[36–39], including specific populations such as people born with missing limbs or patients with epilepsy[40,41]. However, these recruitment benefits depend on the game being perceived as fun by these different groups. Although this is not universally the case (for example, see ref. 42 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[43] in the USA. Some games, such as chess, maintain their broad appeal across cultures to novices and experts alike and now provide enormous and exceptionally rich, diverse datasets of gameplay[44–46].

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 (Fig. 1) collected virtual-navigation data from four million participants in 195 countries, giving insight into why some nations have better navigators[22], how the environment shapes future spatial skill[47] and personalized diagnostics for individuals at genetic risk of Alzheimer's disease[48].

Unlike traditional behavioural experiments, games are unique in that they exist over long timescales. Additionally, because games are enjoyable, participants often revisit games—sometimes over multiple days or even multiple years—and show multiple intermediate milestones of change[49,50]. 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[51]. For example, experienced chess players perceive and remember mid-game positions as larger chunks than novices[52,53]. 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, researchers need to observe people over many more sessions than is practical for in-laboratory testing. To overcome this limitation, Stafford and Dewar[54] worked with the game design studio 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 although 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 that of 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. These can be broadly categorized into two topics: unique challenges of experimental design with games, and data collection and 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. As 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 behaviours 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 participants' prior experiences of the task, which may also affect their performance. These differences can 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[55,56], or creating your own game where direct experimental control is possible[17,22,57]. We detail these considerations further in 'How to use games as a research platform'.

Data collection and analysis

Once a game-based experiment has been designed, collecting and analysing data also presents unique challenges relative to classic psychological tasks. We detail these challenges below.

First, data collection with games can be difficult. 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 (such as virtual laboratories) have been developed, so that researchers have all the usability and functionality they need, while participants can still easily access and participate in the experiments at their own pace[34,58–60].

Once the data have been obtained, they 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 datasets 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 analysing data from games cannot be validated. If such derived measures can also predict everyday behaviour (that is, they have predictive validity), or they produce similar conclusions to established measurements (that is, 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 tasks) and may therefore be difficult to model statistically or computationally[21]. Some researchers have turned to contemporary techniques from machine learning and artificial intelligence to aid with modelling complex game stimuli (such as how to computationally represent pixels on a screen[18,61]). As more complex models, such as multi-layer neural networks or planning algorithms from artificial intelligence, 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[62]. 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. Van Opheusden et al.[21] showed this in a simple two-player game: tree search (tantamount to mentally simulating the 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.

Although we have argued that the data gathered from game-based tasks are more ecologically valid than those from traditional psychology tasks, they will still not perfectly capture natural behaviour (for example, behaviour derived from mobility data[63] or head-mounted cameras[64]). Games lie on a spectrum between traditional laboratory-based tasks (where complete experimental control is possible but ecological validity is low) and natural behaviour (where experimental control is impossible but ecological validity is maximized). Future research will be needed to verify the ecological validity of different games with respect to their natural behavioural counterparts.

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 (for example, by focusing on complexity in different domains such as strategy or input) and how one should set up the game's progression (or curriculum) from level to level. Similarly, because 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 or whether you want to pursue a more portfolio-based approach of using games to study multiple psychological constructs[34,58]. We believe that these choices and considerations should be made with

https://doi.org/10.1038/s41562-024-01878-9

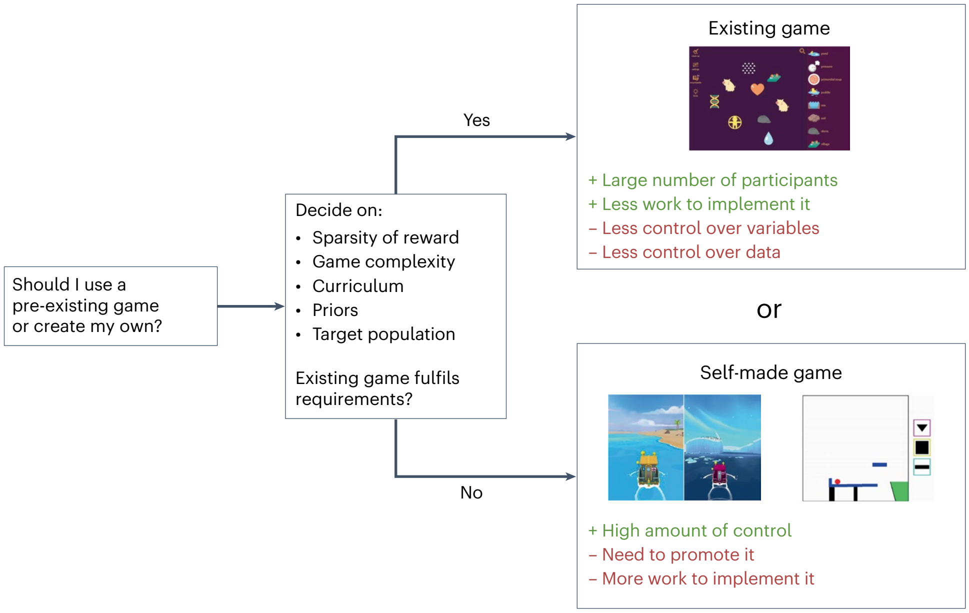

Fig. 2 | Decision criteria, advantages and drawbacks of using existing games and self-made games. Top, *Little Alchemy 2* (ref. 28). Bottom, from left to right, Sea Hero Quest[22] and the Virtual Tools game[17]. Images adapted with permission from ref. 33, Recloak/Jakub Kozioł (existing game) and ref. 17, PNAS (self-made game).

clear hypotheses in mind about the underlying cognitive mechanisms that researchers want to assess[65].

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 (such as chess or *Angry Birds*) (Fig. 2). Although 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[66] or Unity[67], it requires you to think about where to publish the game to recruit enough participants. Using an existing popular game has the advantage of an already existing player base and therefore a substantial amount of data to analyse. However, even if available, these data may not be well suited to answering 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 with 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 (for example, 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[68,69]. Researchers thereby 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 partnerships with game developers, several conferences exist that bring researchers and programmers together (BrainPlay, https://www.cmu.edu/ni/events/brain-play-2022.html; Neurodiversity in Tech, https://pong-center.ucsd.edu/internship/); or ReGame-XR Summer Internship, https://regamexr.sites.northeastern.edu/internship/).

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 ways to overcome these challenges are to (1) work with smaller companies or individual developers who may be more amenable to constructive solution-finding with researchers, (2) work with established games that are not undergoing major development or (3) obtain permission to make a copy of the game for research purposes and host 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 game 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 ref. 42). 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. In this case, 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[70], thus further engaging and diversifying participants[56]. Moreover, researchers can partner with organizations and charities to reach out to different communities to play games.

Finally, we give some advice on how to analyse data collected in games. We suggest storing the ultimately larger datasets 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 datasets, to gauge the relative importance of an effect. Moreover, because 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.

 https://doi.org/10.1038/s41562-024-01878-9

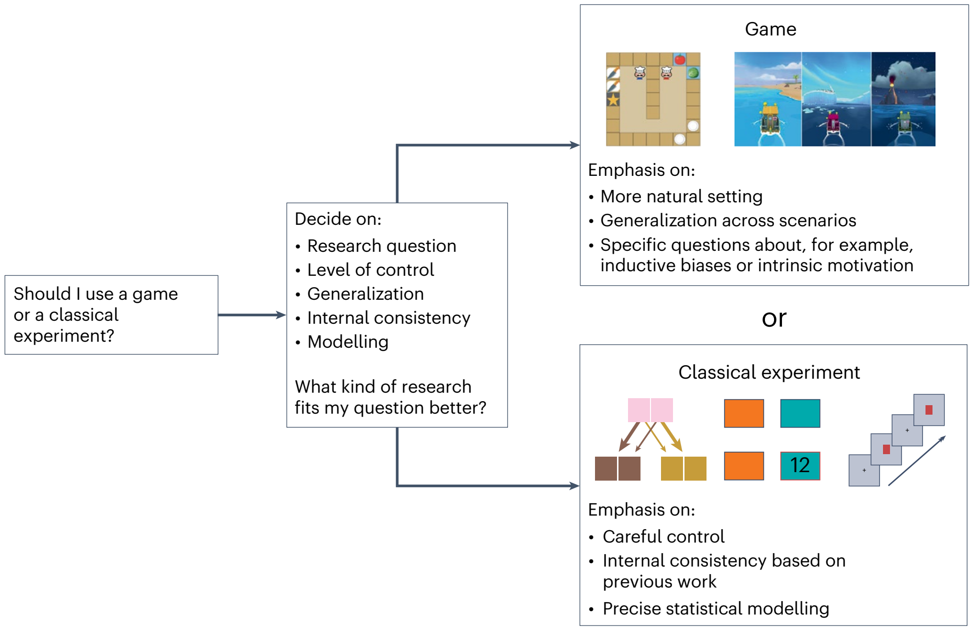

Fig. 3 | Decision criteria to choose between games and classical experiments. Top, from left to right, a cooking game inspired by *OverCooked*[23,24] and Sea Hero Quest[22]. Bottom, from left to right, a two-step decision-making task first introduced by Daw et al.[77], a multi-armed bandit task[79] and an *n*-back memory task[78]. Images of game adapted with permission from ref. 23, Wiley (left) and ref. 22, Glitchers Ltd and Alzheimer's Research UK (right).

Table 1 | Examples of different psychological effects that have been tested using games

Psychological effect	Game	Game properties
Increase of planning depth with expertise[21]	Four-in-a-row	More complex: large state spaces
Cognitive underpinnings of tool use[17]	Virtual Tools	More intuitive: physics-based game dynamics, large action-space complexity
Theory of mind in multi-agent collaboration[23]	*OverCooked*	More complex: multiple levels of hierarchical planning
Usage of empowerment as an exploration strategy[33]	*Little Alchemy 2*	More realistic: requirement of rich semantic priors More enjoyable: testing exploration strategies without explicit goals
Entropy of city street networks linked to future spatial navigation ability[47]	Sea Hero Quest	More enjoyable: supports gathering large-scale datasets

Analysing data from games is usually more difficult than analysing 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 (for example, inverse binomial sampling[71]) and global parameter optimization techniques (for example, Bayesian optimization[72]), respectively, to circumvent these difficulties.

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

In this Perspective, we have tried to distinguish games from classical experiments (Fig. 3). However, the two approaches are not mutually exclusive, and we think that research can profit from combining them. Several of the characteristics differentiating games from experiments are related to the tension between internal and external validity inherent to any empirical study[73]. For example, whereas games allow researchers to test theories in more natural settings, they are less controllable. The reverse is true for experiments. To make statements that are internally consistent but also generalize well across scenarios, we suggest combining games and experiments. Like previous research[34,74,75], 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[34]. We do not believe that games should replace experiments but rather that games and in-laboratory 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 (Table 1). We believe that games have the potential to answer even more questions than we have covered in this Perspective, especially as they relate to intrinsic motivation, holistic cognition, neuroscience and comparative cognition.

A main advantage of games is that they are uniquely positioned to help us answer questions about our innermost motivations. For example, why do we pursue goals despite no obvious extrinsic reward? Several existing popular games emphasize this lack of extrinsic reward. Computer games such as *Minecraft* do not provide specific goals for the player to achieve but rather offer a sandbox within which the player can freely act. What do people decide to do in these scenarios? Do they set goals for themselves to accomplish, and if they do, on what timescale or level of difficulty do these goals usually lie? As classical psychological paradigms rarely provide the opportunity to explore interesting environments without predefined goals, games have the potential to substantially advance this research domain.

More generally, games can steer research in the direction of holistic cognition. Games such as *The Legend of Zelda* require the player to use many different aspects of cognition including planning, exploration and memory. These concepts have traditionally been studied separately from each other, thereby ignoring potential interaction effects. For example, how might the exploration of a game map affect which game mechanics the players remember or how they plan ahead for new enemy encounters? With paradigms such as computer games and their potential to gather very large datasets, cognitive scientists have the opportunity to look at cognition more comprehensively.

Moreover, games are cultural phenomena. They are continuously created and modified for the enjoyment of large groups of people. This requires a sophisticated understanding of not only the initial beliefs and knowledge of potential players but also how those will change as people interact more with the world. For example, most games involve learning new causal rules about how elements in the game will interact (for example, Super Mario mushrooms that affect an avatar's abilities). Looking at which kinds of causal relationships people build into games (for example, which kinds of effects mushrooms might have) could tell us which kinds of causal relationships people expect to be easy or difficult for other people to learn.

Looking beyond human psychology, games may also unlock new research in neuroscience and animal behaviour. For example, recent studies have shown how to set up video-game apparatuses for patients with epilepsy[41], allowing human intracranial recordings to be collected during gameplay. This presents unparalleled possibilities for understanding neural responses to more naturalistic stimuli with tasks that require multiple cognitive systems such as perception, memory, and decision-making. Recent research has shown that primates are similarly able to play reasonably sophisticated games such as *Pac-Man* using compositional and hierarchical strategies[76]. With multiple species playing the same games, there is a unique opportunity for comparative research on how different cognitive systems are recruited for complex, naturalistic cognition at both the neural and behavioural levels. Even with a simple game like *Pac-Man*, these studies could reveal whether people recruit different cognitive systems for decision-making relative to other animals and how specific neural responses may be modulated by task rewards.

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 cognition using 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 if the potential pitfalls associated with the increased complexity found in many games are not minimized. We believe that these pitfalls can be mitigated by making sure that games are designed to test particular hypotheses about human behaviour and by making sure that computational modelling 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 (that is, 'engineered'), and well-designed games can offer such environments.

References

1. Newell, A. et al. *Human Problem Solving* Vol. 104 (Prentice-Hall, 1972).
2. Gobet, F., de Voogt, A. & Retschitzki, J. *Moves in Mind: The Psychology of Board Games* (Psychology Press. 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. *Cognitive Science Society*, https://cognitivesciencesociety.org/wp-content/uploads/2020/06/cogsci20a-W6-doc.pdf (2020).
5. Brändle, F., Allen, K. R., Tenenbaum, J. & Schulz, E. Using games to understand intelligence. *Proc. Ann. Meeting Cogn. Sci. Soc.* **43**, https://escholarship.org/uc/item/17z2q92d (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 (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.* **11**, 2127 (2020).
9. Manzano-León, A. et al. Between level up and game over: a systematic literature review of gamification in education. *Sustainability* https://doi.org/10.3390/su13042247 (2021).
10. Fleming, T. M. et al. Serious games and gamification for mental health: current status and promising directions. *Front. Psychiatry* https://doi.org/10.3389/fpsyt.2016.00215 (2017).
11. Arjoranta, J. How to define games and why we need to. *Comput. Games J.* **8**, 109–120 (2019).
12. Stenros, J. The game definition game: a review. *Games Cult.* **12**, 499–520 (2017).
13. Wittgenstein, L. *Philosophical Investigations: The German Text, with an English Translation by G.E.M. Anscombe, P.M.S. Hacker and Joachim Schulte* 4th edn (Blackwell, 2009).
14. Aarseth, E. in *From Literature to Cultural Literacy* (eds Segal, N. & Koleva, D.) 180–188 (Palgrave Macmillan UK, 2014).
15. Gentner, D. Structure-mapping: a theoretical framework for analogy. *Cogn. Sci.* **7**, 155–170 (1983).
16. Spelke, E. S., Breinlinger, K., Macomber, J. & Jacobson, K. Origins of knowledge. *Psychol. Rev.* **99**, 605–632 (1992).
17. 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. USA* **117**, 29302–29310 (2020).
18. Dubey, R., Agrawal, P., Pathak, D., Griffiths, T. L. & Efros, A. A. Investigating human priors for playing video games. *PMLR* **80**, 1349–1357 (2018).
19. Pouncy, T. & Gershman, S. J. Inductive biases in theory-based reinforcement learning. *Cogn. Psychol.* **138**, 101509 (2022).
20. Tsividis, P. A. et al. Human-level reinforcement learning through theory-based modeling, exploration, and planning. Preprint at https://doi.org/10.48550/arXiv.2107.12544 (2021).
21. van Opheusden, B. et al. Expertise increases planning depth in human gameplay. *Nature* **618**, 1000–1005 (2023).
22. Coutrot, A. et al. Global determinants of navigation ability. *Curr. Biol.* **28**, 2861–2866 (2018).
23. Wu, S. A. et al. Too many cooks: Bayesian inference for coordinating multi-agent collaboration. *Top. Cogn. Sci.* **13**, 414–432 (2021).
24. Carroll, M. et al. On the utility of learning about humans for human–AI coordination. *Proc. 33rd International Conf. Neural Inf. Process. Syst.* (eds Wallach, H. M. et al.) 5174–5185 (Curran Associates, 2019).
25. 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).
26. McCarthy, W. P., Hawkins, R. D., Wang, H., Holdaway, C. & Fan, J. E. Learning to communicate about shared procedural abstractions. Preprint at https://doi.org/10.48550/arXiv.2107.00077 (2021).
27. Jun, E., Hsieh, G. & Reinecke, K. Types of motivation affect study selection, attention, and dropouts in online experiments. *Proc. ACM Hum. Comput. Interact.* **1**, 1–15 (2017).

28. Brändle, F., Binz, M. & Schulz, E. in *The Drive for Knowledge: The Science of Human Information Seeking* (eds. Cogliati Dezzi, I. et al.) 147–168 (Cambridge Univ. Press, 2022).
29. Leonard, J. A., Garcia, A. & Schulz, L. E. How adults' actions, outcomes, and testimony affect preschoolers' persistence. *Child Dev.* **91.4**, 1254–1271 (2020).
30. Chu, J. & Schulz, L. E. Play, curiosity, and cognition. *Annu. Rev. Dev. Psychol.* **2**, 317–343 (2020).
31. Murayama, K. A reward-learning framework of knowledge acquisition: an integrated account of curiosity, interest, and intrinsic–extrinsic rewards. *Psychol. Rev.* **129**, 175–198 (2022).
32. Kosoy, E. et al. Exploring exploration: comparing children with RL agents in unified environments. In Proceedings of "Bridging AI and Cognitive Science Workshop", (2020).
33. Brändle, F., Stocks, L., Tenenbaum, J., Gershman, S. & Schulz, E. Empowerment contributes to exploration behaviour in a creative video game. *Nat. Hum. Behav.* **7**, 1481–1489 (2023).
34. Pedersen, M. K. et al. Measuring cognitive abilities in the wild: validating a population-scale game-based cognitive assessment. *Cogn. Sci.* **47**, e13308 (2023).
35. 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. Res. Methods* **51**, 1782–1803 (2019).
36. 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. Adolesc.* **37**, 1345–1352 (2014).
37. Jansz, J., Avis, C. & Vosmeer, M. Playing the Sims2: an exploration of gender differences in players' motivations and patterns of play. *N. Media Soc.* **12**, 235–251 (2010).
38. 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).
39. Awad, E. et al. The moral machine experiment. *Nature* **563**, 59–64 (2018).
40. Allen, K. R. et al. Lifelong learning of cognitive styles for physical problem-solving: the effect of embodied experience. *Psychonomic Bull. Rev.*, https://doi.org/10.3758/s13423-023-02400-4 (2023).
41. Ashmaig, O. et al. A platform for cognitive monitoring of neurosurgical patients during hospitalization. *Front. Hum. Neurosci.* **15**, 702 (2021).
42. Andrews, G. Gameplay, gender, and socioeconomic status in two American high schools. *E-Learn. Digit. Media* **5**, 199–213 (2008).
43. Duggan, M. *Gaming and Gamers* (Pew Research Center, 2015).
44. Chassy, P. & Gobet, F. Risk taking in adversarial situations: civilization differences in chess experts. *Cognition* **141**, 36–40 (2015).
45. Morsch, F. & Feist, M. Fritz v.12 (Chessbase, 2009).
46. Holdaway, C. & Vul, E. Risk-taking in adversarial games: what can 1 billion online chess games tell us? *Proc. Ann. Meeting Cogn. Sci. Soc.* **43**, https://escholarship.org/uc/item/403764rd (2021).
47. Coutrot, A. et al. Entropy of city street networks linked to future spatial navigation ability. *Nature* **604**, 104–110 (2022).
48. Coughlan, G. et al. Toward personalized cognitive diagnostics of at-genetic-risk Alzheimer's disease. *Proc. Natl Acad. Sci. USA* **116**, 9285–9292 (2019).
49. Chu, Y. & MacGregor, J. N. Human performance on insight problem solving: a review. *J. Probl. Solving* **3**, 6 (2011).
50. Gick, M. L. & Holyoak, K. J. Analogical problem solving. *Cogn. Psychol.* **12**, 306–355 (1980).
51. de Groot, A. D. *Het Denken van den Schaker: Een Experimenteel-Psychologische Studie* (Noord-Hollandsche Uitgevers Maatschappij,1946).
52. Chase, W. G. & Simon, H. A. Perception in chess. *Cogn. Psychol.* **4**, 55–81 (1973).
53. Gobet, F. et al. Chunking mechanisms in human learning. *Trends Cogn. Sci.* **5**, 236–243 (2001).
54. Stafford, T. & Dewar, M. Tracing the trajectory of skill learning with a very large sample of online game players. *Psychol. Sci.* **25**, 511–518 (2014).
55. Eisenberg, I. W. et al. Uncovering the structure of self-regulation through data-driven ontology discovery. *Nat. Commun.* **10**, 2319 (2019).
56. Rafner, J. et al. Digital games for creativity assessment: strengths, weaknesses and opportunities. *Creat. Res. J.* **34**, 28–54 (2022).
57. Schulz, E., Bertram, L., Hofer, M. & Nelson, J. D. Exploring the space of human exploration using Entropy Mastermind. Preprint at *bioRxiv*, https://doi.org/10.1101/540666 (2019).
58. Brown, H. R. et al. Crowdsourcing for cognitive science—the utility of smartphones. *PLoS ONE* **9**, e100662 (2014).
59. Griffiths, T. L. Manifesto for a new (computational) cognitive revolution. *Cognition* **135**, 21–23 (2015).
60. Almaatouq, A. et al. *Scaling Up Experimental Social, Behavioral, and Economic Science* (Computational Social Science Lab at the University of Pennsylvania, 2021).
61. 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).
62. 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).
63. Bongiorno, C. et al. Vector-based pedestrian navigation in cities. *Nat. Comput. Sci.* **1**, 678–685 (2021).
64. Sullivan, J., Mei, M., Perfors, A., Wojcik, E. & Frank, M. C. SAYCam: a large, longitudinal audiovisual dataset recorded from the infant's perspective. *Open Mind* **5**, 20–29 (2021).
65. Rafferty, A. N., Zaharia, M. & Griffiths, T. L. Optimally designing games for behavioural research. *Proc. R. Soc. A* **470**, 20130828 (2014).
66. Pymunk v.6.6.0, https://pymunk.org (Victor Blomqvist, 2023).
67. Juliani, A. et al. Unity: a general platform for intelligent agents. Preprint at https://doi.org/10.48550/arXiv.1809.02627 (2020).
68. Mazarakis, A. & Bräuer, P. Gamification is working, but which one exactly? Results from an experiment with four game design elements. *Int. J. Hum. Comput. Interact.* **39**, 612–627 (2023).
69. Hodges, N. How to gamify cognitive tasks to increase participant engagement.*YouTube*, https://www.youtube.com/watch?v=2TBlII4t3v4&t=2s (Sage, 2023).
70. Silvertown, J. A new dawn for citizen science. *Trends Ecol. Evol.* **24**, 467–471 (2009).
71. van Opheusden, B., Acerbi, L. & Ma, W. J. Unbiased and efficient log-likelihood estimation with inverse binomial sampling. *PLoS Comput. Biol.* **16**, e1008483 (2020).
72. Acerbi, L. & Ma, W. J. Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. *Adv. Neural Inf. Process. Syst.* **30** (2017).
73. Gravetter, F. J. & Forzano, L.-A. B. *Research Methods for the Behavioral Sciences* 4th edn (Wadsworth, 2012).
74. Lewkowicz, D. J. The concept of ecological validity: what are its limitations and is it bad to be invalid? *Infancy* **2**, 437–450 (2001).
75. 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 (2001).
76. Yang, Q. et al. Monkey plays Pac-Man with compositional strategies and hierarchical decision-making. *eLife* **11**, e74500 (2022).

77. 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).
78. Kirchner, W. K. Age differences in short-term retention of rapidly changing information. *J. Exp. Psychol.* **55**, 352–358 (1958).
79. Gershman, S. J. Uncertainty and exploration. *Decision* **6**, 277–286 (2019).
80. Shallice, T. Specific impairments of planning. *Phil. Trans. R. Soc. Lond. B* **298**, 199–209 (1982).
81. Costa-Gomes, M., Crawford, V. P. & Broseta, B. Cognition and behavior in normal-form games: an experimental study. *Econometrica* **69**, 1193–1235 (2001).
82. Schaul, T. A video game description language for model-based or interactive learning. In *2013 IEEE Conference on Computational Intelligence in Games (CIG)* 1–8 (IEEE, 2013).

Acknowledgements

We thank A. A. Kumar and Y. Harel for helpful discussions.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to Eric Schulz.

Peer review information *Nature Human Behaviour* thanks the anonymous reviewers for their contribution to the peer review of this work.