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], 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]University of California San Diego, San Diego, 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, University College London, London, United Kingdom
[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]Center for Humans & Machines, Max Planck Institute for Human Development, 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

Video games are played by over 2 billion people spread across the world population, with both children and adults participating. Games have gained popularity as an avenue for studying cognition. We believe that studying cognition using games can generate progress in psychology and in neuroscience similar to the one that has occurred in artificial intelligence research over the past decades. Using games to understand the mind enables researchers to scale up theories of cognition to more complex settings, reverse-engineer human inductive biases, create experiments that participants want to take part in, and study learning over long time horizons. 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 that this article will lead to a wider use of games as experimental paradigms, elevating the complexity, robustness, and external validity of research on the mind.

Introduction

Progress in psychological and cognitive science has been driven by the development of carefully controllable 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

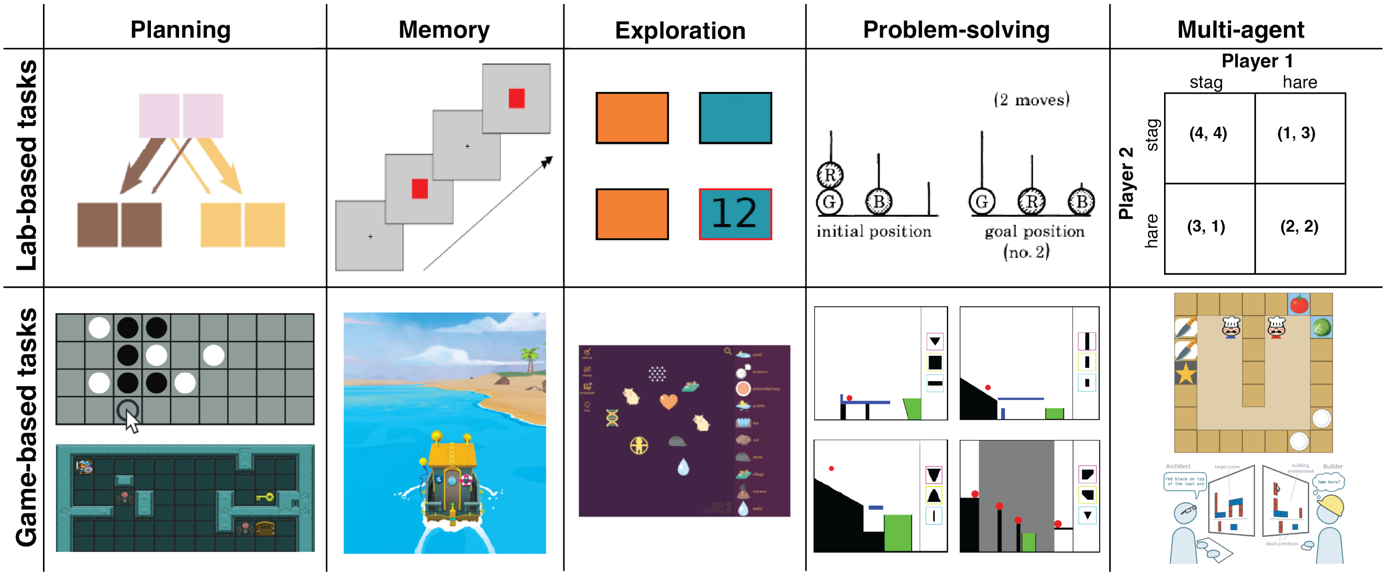

Figure 1. A comparison between classic lab-based tasks (**top**) and games (**bottom**) developed to study different facets of cognition. **Top**: from left to right, a two-step decision making task first introduced by[29], an *n*-back memory task[30], a multi-armed bandit task[31], the Towers of London problem-solving task[32], a matrix-form social coordination task[33]. **Bottom**: from left to right, the "4-in-a-row" game studied by[19] and an example of a programmatically generated video game[25,34], Sea Hero Quest[23], Little Alchemy[27], the Virtual Tools game[26], OverCooked[35,36] and multi-agent construction[37].

inductive biases that support complex action, or what cognitive mechanisms support the intrinsic motivation which compels people to perform tasks (see Fig. 1).

Using complex and richly-structured game-like tasks has been one driver of the sizable increases in the capabilities of artificial systems, from winning a chess tournament[1], to beating the world champion of Go[2], to even winning complex multi-player games like poker[3]. The techniques developed to achieve such abilities have gone on to support very different achievements like protein folding[4] or chip design[5], while some games continue to challenge current artificial systems[6]. We believe that by studying natural intelligence in richly-structured games, we will see similar progress in our understanding of the mind.

Because games are frequently designed to challenge our abilities and capture our interests, they have long been central to the study of the mind[7,8]. Game playing is a popular recreational activity[9] among children and adults, across cultures[10,11], and since ancient times[12]. 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[13], and researchers have learned how to take advantage of the engaging nature of games (through "gamification") for applications in education[14,15] and therapy[16]. 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.

Since providing a formal definition for games is difficult[17], we will characterize games by comparing them to classic experiments using the following design features of games: Games scale up complexity, are intuitive to players[18], engage their players and are able to provide detailed longitudinal data. These design features lead to distinct advantages of using games as a research platform, such as the opportunity to reverse-engineer the set of assumptions that people bring to bear when solving a task, to gather large data sets over much longer time-spans[19,20], to engage a wider diversity of participants[21–23], and to open up new avenues for research such as the study of motivational systems, resource constraints, and richer cognitive priors[24–28].

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 many different types of research, and put forward recommendations on how to best use games in behavioral research. As more daily human experience becomes 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.

Design feature	Advantages	Drawbacks	Solutions
Scaling up using complex settings	• Reveal interactions • Individual differences	• Less controllability	• Black-box techniques • Compare classes of models
Reverse-engineering inductive biases	• More ecological inductive biases	• May be game or creator-specific	• Use popular games • Verify with experiments
Participant engagement	• Research intrinsic motivation • Gather large datasets	• Risk of finding meaningless "statistically significant" results	• Pre-register • Report effect sizes
Longitudinal data	• Analyze learning across time scales • Study curriculum learning & expertise	• Platform needed • Variability between subject's experience	• Use infrastructure like virtual labs • Control for individual experience

Table 1. Overview of different design features of games, their advantages, drawbacks and proposed solutions to overcome the described drawbacks.

Games for scaling up and testing theories

Progress in science is often accompanied by the study of increasingly complex systems, as increased complexity can reveal new causal mechanisms of interaction that are not observed in simpler settings. An important test for any theory is whether it continues to hold in a more complex setting, and if not, which changes are needed to allow it to fit the new data. Games provide an excellent domain in which to test cognitive theories, as they require the interaction of many cognitive domains (for example, perception, motor learning, attention, decision making, etc.), and provide an opportunity to collect a large amount of participant data.

First, scaling cognitive science to larger experiments, with much larger participant pools, can shed light on nuanced differences between individuals[21,38,39], as well as on collective behavior[40]. Games provide an externally validated mechanism for testing theories at larger scales relative to traditional psychological experiments. By attempting to scale theories to more complex domains, it sometimes becomes obvious that simpler, traditional computational models are insufficient[25,27,38,41,42], and new insights are required to explain how people can behave so robustly in these more complex domains. Well-designed games even present powerful opportunities for experimental design at scale as they support a much wider degree of natural variability and randomization than traditional lab experiments. In particular, people can learn a game's core mechanics, but then be exposed to a near limitless set of variations on those mechanics, providing an opportunity for much richer randomization of tasks. In turn, this can support large meta-studies[43] to verify psychological theories at scale.

Second, games can reveal previously unknown interactions between different cognitive processes. While perception, motor learning, attention and decision making are relatively well-understood in isolation, games require people to bring all of them together to play successfully. For example, perception is substantially task-driven while playing Atari games[44], differing motor experiences when growing up with different bodies can affect the way people make abstract cognitive plans in a challenging problem solving game[45], and people do not just decide to explore to reduce their uncertainty, but also consider a rich space of compositional semantic features of objects that will enable them to further explore the game[46].

Although increased complexity can force us to scale up our theories to more realistic tasks, complexity is not always a good thing. Games can be complex in unhelpful ways – sometimes they are designed to attract attention with many superfluous bells and whistles, while their mechanics can also be made to be frustratingly exacting to keep people playing. Whether helpful or unhelpful, an increase of complexity means that games are less controllable. The increased complexity of games can make it difficult to trace back specific changes in human behavior to design changes of the game. This makes it hard to test specific hypotheses and increases the risk of telling "just so" stories about the patterns found in large data sets. For example, whereas in-lab studies might require participants to just choose between a limited set of options, professional video game players can perform up to 500 different actions per minute[47]. In these settings, traditional approaches to computational modeling may become intractable[19]. In some cases, researchers must, therefore, rely on statistics that prefer simple solutions, such strong regularization via Lasso regression, or dimensionality reduction techniques such as neural networks to reduce the complexity of large state representations such as pixel-by-pixel visual input[27,44,48], requiring new ways to think about how scientists should build and interpret computational models. Since more complex models, such as multi-layer neural networks or complex planning algorithms, can be set up in various different ways, we believe that comparing models in games requires a change away from simply comparing one model to another and towards comparing classes of different algorithms[38]. 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,[49] 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. Finally, if new insights about the interactions between different cognitive systems are generated by studying more complex games, in many cases it should be possible to design simplified, more traditional experiments to verify that these interactions hold.

Games as intuitive, engineered environments

Games produce behavior approaching the complexity of the real world, but nonetheless are shaped by the engineering goals of their creators[50]. To be engaging, games must be intuitive by reflecting the assumptions that players bring to the game, and well-calibrated to their motivational and resource constraints by, for example, not requiring millimeter- or millisecond-level precision. In psychology research, these assumptions are often referred to as "inductive biases" – the set of assumptions which 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[51], or object-oriented inductive biases to support perception. For example,[52] showed that infants at approximately four months of age already reason about physical objects according to the concepts of continuity and solidity. The infants reasoned correctly that a ball falling on a surface should not pass through that surface. Computationally, inductive biases are often encoded as priors in Bayesian models[53], as hard-coded structures in logical computational models[54], or as constraints on parametric learning systems.

Games provide a window onto the human inductive biases that support behavior in much more realistic settings than traditional psychological research. For example, while prior work has shown that object-oriented structures are important for human perception, studying behavior in games shows that this is tightly coupled with action. Without object-based world representations, humans struggle to play video games at all – instead performing comparably to machine learning approaches that operate at the level of pixels[48]. Objects alone are not enough; people use *relational* theories to both guide complex, multi-step exploration[25,41], and to communicate knowledge effectively to support further complex action by other individuals[55]. Similarly, prior research points to the importance of intuitive physics as an inductive bias to guide perception[56] and reasoning[57]. However, by studying people's behavior in a more complex physics game,[26] showed that intuitive physics, coupled with object-oriented, relational theories, also supports how people learn to *act* with new tools to solve new problems.

Games are designed to be intuitive to as many players as possible. This is distinct from many classical experiments, which require detailed instructions to understand the task. As a result, games provide an opportunity to more easily investigate cognitive phenomena across cultures[23], 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[35,36] or Codenames[58], where playing successfully critically depends on having *shared* inductive biases with your teammates, or sometimes iterating on communication until such inductive biases are shared[37]. There is significant potential for further exploring how the intuitive nature of games can support such complex behaviors.

Despite being intuitive, games are still engineered by their creators with particular incentive structures in mind. This may dampen generalizability, 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 inductive biases, and may not generalize to other contexts. If using an existing popular game, it may therefore be important to consider complementary games with similar dynamics but different incentives, or by using batteries of games to build cognitive ontologies[59,60].

Games for better participant engagement

Designing experiments that participants want to participate in is important[61]. When completing an experiment at home online, 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 in regular laboratory experiments where participants are compensated for their time – for example, what intrinsically motivates people to explore a new system[27], or to persist in the presence of repeated failure[62]? Curiosity, exploration, fun and play are critical aspects of human cognition[63] which the field currently knows relatively little about, in part because these are difficult concepts to research when rewards are made explicit[64]. 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[65]. Games provide a unique opportunity to learn more about what makes tasks fun, and how people freely behave in tasks which they enjoy.

But even if not studying intrinsic motivation, making tasks fun can lead to an increase in the amount of collected data[21], as well as the diversity of participants[66–69]. Using engaging tasks can even support easier data collection from special populations, such as epilepsy patients[70]. Although different games attract different communities of players (for example, see[71]), as of 2015, regardless of gender or race, people were equally likely to play games, with about 49% of adults playing games occasionally[72] within the United States. Games such as online chess are played across many countries and levels of experience, providing enormous and exceptionally rich, diverse datasets[73,74]. This allows researchers to test theories not only over unseen levels of complexity, but also over many more data points and participant characteristics. For example, the game "Sea Hero Quest" (Fig. 1) collected virtual-navigation data from 4 million participants in 195 countries, giving insight into why some nations have better navigators[23], how the environment shapes future spatial skill[75] and personalized diagnostics for individuals at genetic risk of Alzheimer's disease[76].

Because games are fun, those that are widely available online also present a unique source of existing data to analyze. For example,[39] analyzed millions of chess puzzles to examine players' risk-taking behavior. Researchers can also work with game designers. While game designers may collect data from players in order to improve gameplay and track progress, researchers can analyze the same data with scientific questions in mind. Braendle et al.[27] worked with the creators of a game called "Little Alchemy" to examine sophisticated exploration strategies of human players. While this interaction depends on game creators' willingness to work with academics, academics can also create their own games – either mimicking existing popular ones (such as Mastermind[77]) or building their own games from scratch to study psychological questions at larger scales[24]. Furthermore, we believe that the interaction can also go the other way, where the creators of games can also benefit from scientifically-informed feedback about different design choices.

However, with great data comes great responsibility. Working with large data sets has many potential pitfalls. Beyond practical considerations like where and how to store the data, large data sets bear the risk of finding "statistically significant" results especially in an exploratory setting (e.g., third variable problem or improper corrections for multiple comparisons[78]), and this may not generalize to other games or laboratory tasks. Furthermore, the same game is not equally engaging to all participants. Some participants may return to the game over and over again, while others may play only once. This may exacerbate issues of finding non-generalizable "statistically significant" results as a consequence of non-representative data. It is therefore especially important to derive hypotheses theoretically and possibly even pre-register analyses for game data sets, and also report effect sizes to avoid such pitfalls[75].

Games for studying learning across time-scales

Unlike traditional behavioral experiments, games are unique in that they exist over long time scales, and participants can revisit games over multiple days, or even multiple years, and show multiple intermediate milestones of change[79,80]. Examining such longitudinal data can reveal effects of learning that are otherwise invisible in short periods of simple laboratory tasks[81,82].

At the most extreme end of this scale, games are used to study the acquisition of expertise and related representational and strategy changes, which can take years to develop[83]. For example in chess, experienced players perceive and remember mid-game positions as larger chunks than novices[84,85]. With increased computational power, more recent work has been able to precisely model the acquisition of such expertise, suggesting that expertise is driven by people learning to search through the problem space more deeply[49], or by acquiring increasingly abstract search strategies to reduce the time spent looking for solutions to a problem[37]. Further investigating the role of curricula, or how games are designed to introduce increasingly complex concepts, could support applied psychological research in education.

Games themselves can have long-term learning dynamics, with many including "curricula" which increases game complexity gradually. These can be intrinsically generated, where players explore more of their environment as they progress, or extrinsically generated, where the game explicitly introduces new abilities, levels, or areas to discover over time. In either case, these curricula are shaped by the game designers, which enables some adaptive control over difficulty[86], and better comparisons across many different conditions and capabilities for both within and between-subjects designs.

Furthermore, longitudinal data collection comes with some unique challenges. First, many standard online experimental platforms do not easily support longitudinal data collection. Instead, a special infrastructure – like virtual labs – may be essential, 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[22,24,87]. Second, when gathering data over a longer time-scale, there can be more variability between subjects' experiences that quickly accumulates. For example, 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 discussing the game in online communities. This missing control is an intrinsic complication of longitudinal studies using games.

How to use games as a research platform

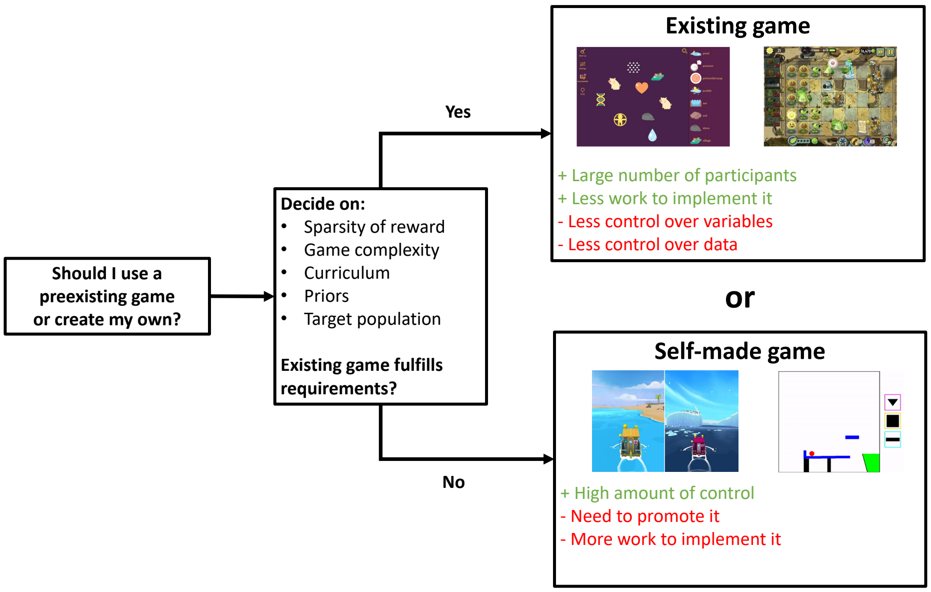

Figure 2. Decision criteria, advantages and drawbacks of using existing games and self-made games. **Top**: from left to right: Little Alchemy 2[27], Use your Brainz[88]. **Bottom**: from left to right: Sea Hero Quest[23], the Virtual Tools game[26]

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. First, where is your game going to come from? Broadly, there are two options: you can use a pre-existing game (like Chess, Angry Birds, etc.) by either working with a game company or using an open source game, or create your own game (see Fig. 2). While creating your own game gives you more control over specific parameters of the game, 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. In this case, we recommend creating your own game.

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 or if you want to pursue a more portfolio-based approach of using games to study multiple psychological constructs[22,87]. 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[89].

With regards to attracting a target population, 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 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[71]). 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, to put the game on different app stores, or –if necessary– to work with a game company directly. Publicly announcing that results are going to be used for research purposes may also call citizen scientists into action[90], thus further engaging and diversifying participants[60]. Moreover, researchers can also partner with organizations and charities to reach out to different communities to play games.

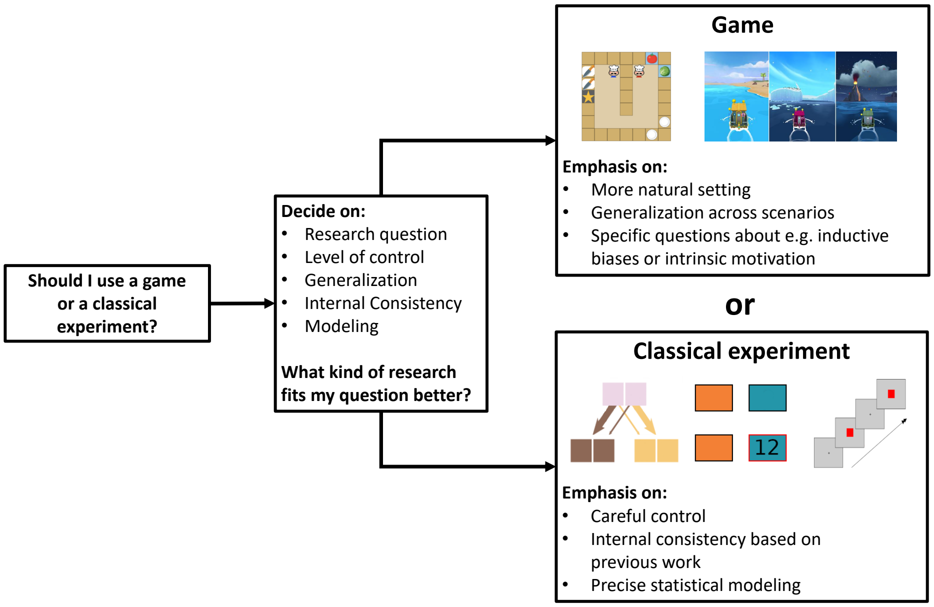

Figure 3. Decision criteria to choose between games and classical experiments. **Top**: from left to right: OverCooked[35,36], Sea Hero Quest[23]. **Bottom**: from left to right: a two-step decision making task first introduced by[29], a multi-armed bandit task[31], an *n*-back memory task[30]

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[42]) and global parameter optimization techniques (e.g., Bayesian optimization[91]), respectively, to circumvent these difficulties.

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[92]. 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[87,93,94], 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[87]. We do not believe that games should replace experiments but rather that games and in-lab experiments will best work in tandem.

Conclusion

We have argued that truly understanding the mind requires a paradigm shift similar to what has happened in machine learning, 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. Campbell, M., Hoane Jr, A. J. & Hsu, F.-h. Deep blue. *Artif. intelligence* **134**, 57–83 (2002).

2. Silver, D. *et al.* Mastering the game of go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).

3. Brown, N. & Sandholm, T. Superhuman ai for multiplayer poker. *Science* **365**, 885–890 (2019).

4. Jumper, J. *et al.* Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).

5. Mirhoseini, A. *et al.* A graph placement methodology for fast chip design. *Nature* **594**, 207–212, DOI: 10.1038/s41586-021-03544-w (2021).

6. Wang, J. *et al.* Alchemy: A benchmark and analysis toolkit for meta-reinforcement learning agents. In Vanschoren, J. & Yeung, S. (eds.) *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, vol. 1 (2021).

7. Newell, A., Simon, H. A. *et al.* *Human problem solving*, vol. 104 (Prentice-hall Englewood Cliffs, NJ, 1972).

8. Gobet, F., de Voogt, A. & Retschitzki, J. *Moves in Mind: The Psychology of Board Games* (2004).

9. Gough, C. Number of gamers worldwide 2021. **13**, 2020 (2019).

10. Suchow, J. W., Griffiths, T. & Hartshorne, J. K. Workshop on scaling cognitive science. In *CogSci* (2020).

11. 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).

12. DePaulis, T. Board games before ur? *Board Game Stud. J.* **14**, 127–144 (2020).

13. 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).

14. 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).

15. 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).

16. 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).

17. Arjoranta, J. How to define games and why we need to. *The Comput. Games J.* **8**, 109–120 (2019).

18. Do, Q., Kane, G. A., McGuire, J. T. & Scott, B. B. Assessing evidence accumulation and rule learning in humans with an online game. *bioRxiv* (2022).

19. van Opheusden, B. *et al.* Revealing the impact of expertise on human planning with a two-player board game. *PsyArXiv* (2021).

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

21. 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).

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

23. Coutrot, A. *et al.* Global determinants of navigation ability. *Curr. Biol.* **28**, 2861–2866 (2018).

24. Almaatouq, A. *et al.* Scaling up experimental social, behavioral, and economic science. 1–8 (2021).

25. Tsividis, P. A. *et al.* Human-level reinforcement learning through theory-based modeling, exploration, and planning. *arXiv preprint arXiv:2107.12544* (2021).

26. 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).

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

28. Rutledge, R. B. *et al.* Association of neural and emotional impacts of reward prediction errors with major depression. *JAMA psychiatry* **74**, 790–797 (2017).

29. 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).

30. Kirchner, W. K. Age differences in short-term retention of rapidly changing information. *J. experimental psychology* **55**, 352 (1958).

31. Gershman, S. J. Uncertainty and exploration. *Decision* **6**, 277 (2019).

32. Shallice, T. Specific impairments of planning. *Philos. Transactions Royal Soc. London. B, Biol. Sci.* **298**, 199–209 (1982).

33. Costa-Gomes, M., Crawford, V. P. & Broseta, B. Cognition and behavior in normal-form games: An experimental study. *Econometrica* **69**, 1193–1235 (2001).

34. 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).

35. Wang, R. E. *et al.* Too many cooks: Coordinating multi-agent collaboration through inverse planning. *CoRR* **abs/2003.11778** (2020).

36. Carroll, M. *et al.* On the utility of learning about humans for human-ai coordination. *Adv. neural information processing systems* **32** (2019).

37. 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).

38. 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).

39. 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).

40. Chen, M. G. Communication, coordination, and camaraderie in world of warcraft. *Games Cult.* **4**, 47–73, DOI: 10.1177/1555412008325478 (2009).

41. Pouncy, T. & Gershman, S. J. Inductive biases in theory-based reinforcement learning. (2022).

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

43. Baribault, B. *et al.* Metastudies for robust tests of theory. *Proc. Natl. Acad. Sci.* **115**, 2607–2612 (2018).

44. 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).

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

46. 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).

47. Vinyals, O. *et al.* Starcraft ii: A new challenge for reinforcement learning. *arXiv preprint* (2017).

48. Dubey, R., Agrawal, P., Pathak, D., Griffiths, T. L. & Efros, A. A. Investigating human priors for playing video games. *CoRR* **abs/1802.10217** (2018).

49. Van Opheusden, B., Galbiati, G., Bnaya, Z., Li, Y. & Ma, W. J. A computational model for decision tree search. In *CogSci* (2017).

50. Salganik, M. J. *Bit by bit: Social research in the digital age* (Princeton University Press, 2019).

51. Gentner, D. Structure-mapping: A theoretical framework for analogy. *Cogn. science* **7**, 155–170 (1983).

52. Spelke, E. S., Breinlinger, K., Macomber, J. & Jacobson, K. Origins of knowledge. *Psychol. review* **99**, 605 (1992).

53. Griffiths, T. L. Bayesian models as tools for exploring inductive biases. In Banich, M. T. & Caccamise, D. (eds.) *Generalization of knowledge: Multidisciplinary perspectives*, 135–156 (Psychology Press, 2010).

54. Forbus, K. D. Qualitative process theory. *Artif. intelligence* **24**, 85–168 (1984).

55. Tessler, M. H., Tsividis, P. A., Madeano, J., Harper, B. & Tenenbaum, J. B. Growing knowledge culturally across generations to solve novel, complex tasks. *arXiv preprint arXiv:2107.13377* (2021).

56. Battaglia, P. W., Hamrick, J. B. & Tenenbaum, J. B. Simulation as an engine of physical scene understanding. *Proc. Natl. Acad. Sci.* **110**, 18327–18332 (2013).

57. Hamrick, J. B., Battaglia, P. W., Griffiths, T. L. & Tenenbaum, J. B. Inferring mass in complex scenes by mental simulation. *Cognition* **157**, 61–76 (2016).

58. 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).

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

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

61. 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).

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

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

64. 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).

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

66. 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).

67. 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).

68. 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).

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

70. 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).

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

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

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

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

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

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

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

78. Gelman, A. & Stern, H. The difference between "significant" and "not significant" is not itself statistically significant. *The Am. Stat.* **60**, 328–331 (2006).

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

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

81. Listman, J. B., Tsay, J. S., Kim, H. E., Mackey, W. E. & Heeger, D. J. Long-term motor learning in the "wild" with high volume video game data. *Front. Hum. Neurosci.* DOI: 10.3389/fnhum.2021.777779 (2021).

82. 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.

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

84. 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.

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

86. Plass, J. L., Homer, B. D., Pawar, S., Brenner, C. & MacNamara, A. P. The effect of adaptive difficulty adjustment on the effectiveness of a game to develop executive function skills for learners of different ages. *Cogn. Dev.* **49**, 56–67 (2019).

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

88. 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).

89. 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).

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

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

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

93. 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.

94. 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.