Abstract

The Skinner box has provided a standardized method of conducting experiments on operant behavior throughout the history of behavior analysis. As technology has advanced, these operant chambers have become increasingly complex to allow for the study of new stimuli, behavior, and outcomes. The present paper takes this variability one step farther by advocating for the use of video game engines in the study of operant behavior in humans. Game engines provide high levels of flexibility, control, and realism as well as continuous behavioral tracking, dynamic stimulus presentation, and complex reinforcement schedules that greatly expands the range of research questions that can be addressed. Importantly, the potential for increasing stimulus, response, and outcome variability provides the basis for assessing and maximizing the generalizability of operant and related principles. The paper illustrates the use of video game engines to study causal inference, delay discounting, loot boxes, foraging, and multiplayer dynamics. Adopting game engines in behavioral research not only expands the scope of behavior analysis but also increases its relevance to real-world behavior, offering a promising path forward for innovation.

Keywords: Video game, Skinner box, behavioral variability, generalizability

Game Engines as the New "Virtual" Skinner Box

The ubiquitous Skinner box (or operant conditioning chamber) and its variants have been in use since at least 1930 (Skinner, 1930). In its creation and evolution, the box allowed the experimenter to present discriminative stimuli (turn on a light, play a sound over a speaker), record one of a discrete set of behaviors (press on the left lever, pull on a chain), and present a reward (a cup of grain or a food pellet) in a defined location. Over the years, its sophistication grew but only modestly. Some modifications included the ability to present a color photograph either via a slide deck, a computer monitor, or a tablet (Atak & Nevin, 1983; Wasserman et al., 1988), to present videos on a monitor (Shimizu, 1998; Young et al., 2006), and to register nose pokes and key pecks on a touch-sensitive screen (Pineño, 2014; Ushitani & Fujita, 2005) or to a defined location (Zeigler & Feldstein, 1971). Recent innovations have included the rich recording of behavior performed within a Skinner box by, for example, taking videos of behavior that are later scored by humans or algorithms (Blackwell et al., 2018; Burgos-Robles et al., 2017; Wittek et al., 2023).

A historical benefit of the Skinner box was standardization. Architecting chambers with similar dimensions, manipulanda, lights, and feeders increased the anticipated generalizability of results across experiments and laboratories. Furthermore, creating sparse environments reduced the influence of factors that create unwanted sources of variance or that can even introduce confounds. As the aforementioned innovations reveal, however, behavioral and neural scientists' research questions necessitated revising the box to examine generalization to new stimuli, response requirements, and outcomes. Given the importance of reproducibility across laboratories and environments, greater attention to the factors that influence reproducibility is necessary. Note that we are making a key distinction between replicability and reproducibility in which the former focuses on obtaining the same result under the same conditions whereas the latter emphasizes the generalizability of a result across situations

Furthermore, the increasing application of basic behavioral paradigms and theoretical frameworks to human subjects divulged a new issue – unlike rats or pigeons in a laboratory environment, humans are not captive and are less likely to repeatedly return for many days on end without significant incentives (e.g., Baron & Galizio, 1976; Madden & Perone, 2003). As a result, a number of the studies of the experimental analysis of behavior in humans involve short duration experiments, often less than an hour of behavior from each subject (e.g., Klapes et al., 2020; Macaskill et al., 2023; Reed et al., 2018). If there is any hope of consistently collecting behavior from people for much longer periods of time, it will prove necessary to create more engaging tasks than the ones adapted from rat and pigeon studies.

In this manuscript, we will consider the traditional benefits of the use of the Skinner box but also its drawbacks. To begin, we consider the classic experimental dilemma between experimental control and generalizability. We will then progress to an alternative approach to the study of human behavior using prebuilt frameworks for developing video games (*game engines*) that empowers the scientist to investigate the generalizability of behavioral principles to more realistic environments with a greater range of stimuli, responses, and outcomes. The use of game engines will also reveal other benefits including the collection of highly detailed behavioral data, greater flexibility in controlling environmental dynamics, and the ease of studying social interactions in multiplayer environments. After examining the challenges that emerge in the use of game engines, we will present examples of gaming environments used to study five types of research questions within our own laboratory. Throughout, we hope to demonstrate the potential for theoretical development, increased creativity, greater enthusiasm for behavioral research, and the vast horizons of research questions that such a move can address.

Before we continue, we wish to stress that this is not a treatise on gamification. Our focus is on the role of experimental variability, generalization across tasks and environments, subject engagement, and behavioral assessment. These factors influence the research questions that can be asked, the quality of the data, and reproducibility within and across contexts. We believe that faster scientific progress in understanding reinforcement principles can be achieved through the use of game engines.

The Role of Experimental Variability

The Skinner box provided some important constraints on behavioral research. The principal advantage of a single relatively uniform platform is that it holds spurious variation constant thus reducing “noise” in the data. This uniformity reduces the likelihood that uncontrolled factors are at the root of differences in results across experiments and laboratories. Of course, some variation still exists because different laboratories vary in rat strains, bedding, lighting, box dimensions, feed, experimenters, etc. But, to the degree that the laboratories use similar conditions, then results should be reproduced across testing sites. There is a concern, however, that this lack of variation could produce results that are specific to a narrow set of conditions thus giving rise to a failure to reproduce an effect when task characteristics change even modestly – in other words, results produced under high levels of standardization may not reproduce well across task variations (Richter et al., 2010; Richter et al., 2009).

Achieving the goal of precise experimental control while also ensuring that the results generalize to new populations and situations creates the experimenter’s dilemma – maximizing internal validity (control) undermines external validity (generalizability) (Campbell, 1957; Lin et al., 2021). In the typical experiment, internal validity is paramount and researchers aim to test theories by using carefully controlled designs. The various threats to internal validity (e.g., non-random assignment to conditions, experimental confounds, differential attrition) invalidate the inferences being tested. It is important to note, however, that lack of control does not automatically create a confound. Random noise (error variance) is present in every experiment and, in a well-designed study, is not systematically related to the independent variables being tested. For example, the use of random assignment of subjects to conditions minimizes systematic condition differences in pre-experimental histories, sex, weight, strain, etc., while still allowing for variability in those factors. Indeed, the National Institutes of Health instituted policies in 2014 to encourage the use of male and female animal research subjects because findings involving only male animals do not always generalize to females (Clayton & Collins, 2014).

In order to establish generalizability, scientists can identify the conditions under which an independent variable's effect might vary in strength, direction, or functional form. It is the rare psychologist, however, who would attempt to examine whether an effect generalizes to a new bedding, day/night cycle, lever manufacturer, or feed. Behavioral sensitivity to these apparently inconsequential variables does occur and undermines reproducibility when they are tightly controlled.

For example, in a study by Richter et al. (2009), the authors compared the effect of standardization against heterogenization. Standardization involved holding a range of factors constant. Heterogenization introduced carefully selected variation across two ages, enrichment levels, cage sizes, test lights, times of day, sound levels, experimenter familiarity, and pretest handling. Two of these factors were chosen and systematically varied in four separate conditions (heterogenization) whereas the same two factors were held constant in matching controls (standardization); the non-selected six factors were held constant in both the heterogenization and standardization conditions. What Richter et al. discovered was that standardization produced larger effects than did heterogenization but with greater variation across replicates – in other words, standardization was more likely to find an effect (or a larger effect) but with a greater false positive rate when applied to other contexts (e.g., to a study involving a different cage size and lighting). It is thus easy to understand why standardization is the norm because it increases the initial likelihood of statistical significance, but this sensitivity comes at the cost of poorer generalizability across experimental variation in purportedly irrelevant variables like sound level or time of day – standardization can reveal an effect or effect size that may not be reproduced across conditions.

It is, of course, critical that any added or existing experimental variation is not confounded with the predictor variables of interest. The concept of randomly assigning subjects to groups can be extended to randomly assigning contexts to conditions. For example, we can design a study in which rats start in random locations in a water maze, the order of conditions randomly varies across blocks, or the amount of reward randomly varies across trials in a Skinner box. Such variation is sometimes present, sometimes systematic (e.g., through the use of counterbalancing), but often not considered. In all cases, systematically introducing variability may be critical to predicting when replicability is expected and under what conditions.

Thus, there is one behavior analytic principle that we are asking to be relaxed – the precise control of the testing environment. B.F. Skinner (1957) placed a lot of emphasis on control, a laudable scientific goal. To achieve that goal, he advocated for intense study of an organism, often the pigeon, which he hoped to be representative of other organisms but (by his own admission) was convenient. He also focused on "a bit of behavior—not for any intrinsic or dramatic interest it may have, but because it is easily observed, affects the environment in such a way that it can be easily recorded, and for reasons to be noted subsequently, may be repeated many times without fatigue" (p. 344). Finally, he constructed the operant chamber to create an environment that can be well controlled. Many of these choices were made out of convenience, a starting place that generated an enormous amount of research and enduring behavioral principles. We suggest, however, a need to accelerate innovation in the field of behavior analysis by increasing our understanding of how contextual variables and individual characteristics impact their operation.

Greater Control of Variability using Game Engines

A classic Skinner box typically provides for one response type (e.g., a lever press), a limited set of stimuli (colored lights and tones), and one outcome (depending on box design – pellets, grain, milk). This limited variation is not only a byproduct of controlling for extraneous variables but also due to technical limitations and the commercialization of apparatus production. Designing boxes that can deliver different reward types or that allow variation in response type (or even just the required response pressure) can significantly increase costs. The introduction of monitors with touch-sensitive screens to present stimuli greatly expanded the horizons of research questions involving stimulus characteristics, but limitations persisted. Thus, even animal labs are recognizing the need to move beyond the limitations of the traditional Skinner box despite the increased costs of overcoming the technical challenges in such a move.

In a typical operant paradigm, the subject's behavior determines what is experienced – what is seen, heard, and rewarded are all contingent on behavior. An understanding of the nature of that contingency, however, is limited by the discrete nature of most operant designs. For example, a bar press may turn on a light or cause a food pellet to be delivered. There are, however, a plethora of contingencies based on continua. As a subject moves its head, the intensity of a light may change as well as its location and size on the retina. A Skinner box is not generally equipped to track gaze location, the distance between the subject and the light source, or similar variables that vary as a function of the subject's behavior. Game engines, however, allow for the measurement of a wide range of continuous behaviors, permitting a more proper treatment of such contingencies.

To appreciate the difference between environments that strongly limit what can be learned about continuous contingencies and those that do not, consider an experiment assessing causal judgment (Michotte, 1946/1963). In a demonstration producing strong causality judgments, one object moved toward another object until the two were contiguous, the first object stopped, and the second object immediately moved away from the stationary first object along the same trajectory (Michotte's classic "launching effect"). In a classic design involving limited observation, a video is recorded (or simulated) showing a single perspective of the launch. A scientist could generate variations on this video to simulate observing it from different distances and angles, but the scientist must decide in advance which ones to generate. In an alternative design (one created using a game engine), a subject can choose their observational perspective, and their choices can reveal their preferences. Furthermore, a game engine can allow the subject to control the behavior of the first object by changing its direction and speed to determine the effects of this variation on the second object's behavior. The game engine approach expands the range of environments tested while also providing additional information about the subject's preferred viewpoints and launching variations. This greater range of behavioral and stimulus variation may either be of concern or considered a feature – a feature because this variation can be studied as a source of outcome variance or because it expands the contexts in which the phenomena were studied.

So, what precisely is a game engine? Game engines are software development environments that include libraries of functionality and prebuilt environmental objects that are helpful in producing a video game. A video game is nothing more than a virtual environment in which a physics engine determines how objects behave (e.g., when they collide or when an object falls off a ledge), allowable actions and their consequences are defined (e.g., firing an arrow versus throwing one), the transmission of sound and light (e.g., sounds are quieter when their source is farther away), and the triggers that produce environmental change (e.g., entering a room turns on a light, plays a sound, or sets off a trap). Existing code libraries enable programmers to create an environment without the need to learn physics equations, and many game engines have vast repositories of user-created assets created by other users including prebuilt environments, avatars and characters, tools like armor, transportation, or weapons, behavioral repertoires for avatars (e.g., jumping, sprinting, climbing), and more.

Thus, game engines make it more tractable for a behavioral scientist to incorporate a vast array of stimuli, responses, and outcomes, thus increasing both the range of possible research questions and the variation in environments in which questions are addressed. They also permit systematic testing of stimulus configurations, including variations in stimulus strength, timing, or combinations that are rare in nature or otherwise difficult to create. Stimuli can easily encompass multiple features, several modalities, dynamically changing stimuli, and stimulus interactions. Information can also be presented via a heads-up display, providing cues about factors that may be difficult to monitor or recall independently, such as health, resources, time, or task progression. This makes it possible to reduce memory-related confounds by ensuring participants have access to information such as instructions, current scores, or completed tasks. Response types can include the various responses produced by a game controller including choices and decision times, as well as avatar movement, mouse movements, and avatar direction of gaze. Outcomes can include diverse game-useful commodities of various amounts (currency, experience points that contribute to progression, and additional ammunition or tools), and achievements that can trigger new abilities, relative rankings compared to other players, and points. Penalties can likewise take many forms, including loss of health, items, or abilities, or decreases in ranking or points.

The advantage of incorporating a greater range of stimuli, responses, and outcomes is that that reinforcement principles studied in one domain may not translate unchanged to new domains. It is already well-understood that there are individual differences between subjects in their sensitivity to stimulus characteristics (e.g., sound and light intensity, Greem & Luce, 1974; Stevens, 1971), their rate of responding (Harzem, 1984), and the motivational impact of outcome characteristics (e.g., delay and magnitude, Jarmolowicz et al., 2020; Steele et al., 2019). It is also conceivable that some reinforcement schedules (e.g., variable interval) are more effective at controlling a discrete behavior like firing a weapon than a continuous behavior like running. Exploring a wider array of environments will reveal such differences. Thus, we agree with Buskist et al. (1985, p. 72) that “One of the more critical objectives of the experimental analysis of *human* behavior is to enhance the breadth of research topics addressed by its practitioners” (emphasis added).

Generalization to More Realistic Environments

The Skinner box provided a nicely controlled environment for the study of principles of learning, but the typical box does not resemble the natural environment of the species studied.

The elevation of internal validity over external validity has costs as well as benefits. A lot of applied research hinges on knowing whether the discovered principles apply across situations, and applied behavior analysts have led the way in extending behavior principles to environments where behavioral change is needed by having “shifted their application exclusively to what the group called ‘socially important behaviors’ in those behaviors’ real-life Skinner boxes” (Baer et al., 1987, p. 313).

Many environments can benefit from the application of behavioral principles, and game environments open the door to examining generalization to new domains. Too often, researchers in other research domains have rediscovered behavioral principles because they did not see the relevance of the study of bar presses, tones, and lights to decisions involving investments, battlefields, diets, or taking an exam. In our opinion, behavior analysis has ceded too much territory in the learning domain to other specialties like cognitive training, reinforcement learning algorithms, coaching, classroom learning, and physical training. If those fields were applying principles originally discovered using Skinner boxes, we would have less concern. Unfortunately, most of the researchers studying decision making and learning are ignorant of many of the behavioral principles so well understood in behavior analysis. Although this ignorance may reflect pedagogical approaches currently in vogue in psychology (i.e., not behaviorism), at least some of this ignorance is willful when researchers doubt the relevance of work on rats and pigeons in limited environments to addressing the real problems that they are trying to solve (Holleman et al., 2020; Maes et al., 2015; Mills, 1899).

For example, gambling researchers and game developers are aggressively seeking methods to get players to continue to play (Griffiths, 1993; Newall et al., 2025; Toneatto et al., 1997). Gambling researchers sometimes have awareness of reinforcement principles (e.g., James et al., 2016; Ramnerö et al., 2019) as do some game developers (but see Epifânio & da Silva, 2023; Yee, 2014). A lack of awareness, however, does not prevent game designers from either rediscovering principles pioneered in behavior analysis or from discovering new principles less-informed by basic science (e.g., Juul, 2012; Vorderer et al., 2003; Yee, 2006). Indeed, there are undoubtedly things that behavior analysis can learn from game developers (e.g., Allen et al., 2024; Bouvier et al., 2014; Harrison & Roberts, 2015; Hendrix et al., 2018; Morford et al., 2014).

Basic behavioral research must be careful not to spend too much time studying easy, tractable problems. Staying in the Skinnerian sandbox undermines the creation of an empirical base that would drive theoretical development. We are not arguing that the sandbox must be left behind because there is nothing new to learn. Rather, we are suggesting that we expand our theoretical and empirical horizons as well as extend our work to more realistic environments in order to help others see the relevance of our work.

Decreasing the gap between laboratory studies and the context to which their results will be applied will likely have greater appeal to funding agencies, other scientists, and the general public (Buskist et al., 1985). Although there are mechanisms to fund basic science, the average citizen and politician often struggles to see the relevance of basic research and its economic benefits (Nelson, 1959; Pavitt, 1991). When basic research on impulsivity is tied to drug abuse, battlefield decisions, or investing, it is easier for the public to appreciate the utility of the work being supported. We are definitely not encouraging an abandonment of basic research. Rather, many of the choices that we make regarding required responses, stimuli, and rewards have been choices of convenience, and the simple decision to champion variability ensures that our results generalize to a greater variety of contexts while also making it easier for the public to appreciate the full gamut of implications of our work. Given the economic benefit of basic research and its contribution to innovation (Azoulay et al., 2019), increasing public support for basic behavioral research by taking small steps toward real world settings should not be too much to ask.

More Engaging Tasks Lead to More Engaged Students and Scientists

We believe that the use of video game platforms to study behavior may also have a secondary benefit – engaging students and scientists. By "engagement" we mean the ability to initiate a desired behavior and to sustain it. Too much of the classic learning and reinforcement research and the associated journal articles and courses are at risk of turning people away. The field's jargon is deemed as unpleasant (Critchfield et al., 2017), and the primary subjects are often rats, pigeons, and mice which will seem less relevant to understanding people's behavior. Indeed, there are many published failures to generalize results from animal (pre-clinical) models to humans (Brubaker & Lauffenburger, 2020; Henderson et al., 2013; Pound & Ritskes-Hoitinga, 2018), so it is not surprising that human subject research in the field's flagship journal, *Journal of the Experimental Analysis of Behavior*, has steadily increased from a small minority of publications in the early years (Saville et al., 2002) to 47% by 2016-2018 (Galizio, 2020).

When appropriate, implementing behavioral interventions in the target species – humans – using more interesting environments may serve to attract researchers and students and increase the penetration of behavior analysis into more laboratories. Although roughly 68% of U.S. four-year psychology departments require undergraduates to participate in research (Flynn & Rocheleau, 2024), many Skinnerian-like experiments, including prior experiments from our own laboratory, are overly tedious and repetitive leading to poor engagement. If we want to inspire the next generation of behavioral analysts and encourage more undergraduates to get involved in research, then creating engaging experiments may be our best (and sometimes our only) chance.

By using video game engines to craft richer, more interactive tasks, we may encourage more undergraduate students to participate in our studies and potentially attract more students to become research assistants or even pursue careers in behavioral analysis. Although this approach suggests that marketing of behavior analysis is necessary (Bailey, 1991; Freedman, 2016; Smith, 2016), scientists who want to increase the impact of their work, or their field, must be intentional to achieve this goal.

We believe that the experimental analysis of behavior is a critically important research paradigm, but its influence has waned over the decades. Behavior analysis needs more innovation if the field is going to increase its relevance and thus its impact by attracting and retaining students, scientists, and practitioners. Video games play a major role in people's everyday lives (Entertainment Software Association, 2025) with 60% of American adults and 84% of children reporting playing video games at least one hour per week in 2024 – that represents a lot of behavior that can be recorded at a high resolution.

In our experience, undergraduate and graduate students find the idea of using games in experimental psychology to be appealing. By drawing on their own experiences, they readily see the applicability of behavioral principles to good game design. In contrast, the Skinner box and other standardized platforms and tasks can lead to "focused efforts to study toy problems in well-contained scenarios" (Dubova et al., 2025). Although such focused efforts can lead to a steady stream of publications that advance a scientist's career by providing clear outcomes, this churn can stifle creativity in task design, theory development, and the solution of applied problems. We are confident that the greatly increased horizons of video game engines will spur innovation, and behavior analysts can lead the way. Behavior analysis emphasizes within-subject designs in which an organism is studied for prolonged periods; these designs produce long bouts of behavior like those produced in game play. Furthermore, the field's focus on operant behavior is facilitated by using game engines due to their ability to finely tune the stimulus conditions, to leverage a much wider range of behaviors, and to use outcomes that can sustain prolonged behavior without significant satiation. Satiation may also be avoided game by introducing variation through regular updates, patches, etc., that change the contingencies or introduce new game elements.

Richer Behavioral Assessment

One surprising benefit of our use of video game platforms for data collection is the host of behavioral variables that can be easily collected at low cost (Pedersen et al., 2023; Rafner et al., 2022). In addition to the standard responses and latencies, we can track the moment-by-moment coordinates of the avatar, the orientation of the avatar (effectively equivalent to gaze direction, determining what content is available on-screen and thus available for visual processing by the participant), allowable avatar actions like jumping or reaching or opening a door, inter alia. The limitations of bar pressing, key pecking, and chain pulling become a thing of the past. Indeed, there is so much data that can be collected that it can become overwhelming. In our own use of video game engines, we often record more data than we need for the purposes of an experiment. Given that the cost of collecting these extra variables is trivial, it has proven useful to record key variables that might prove useful in subsequent studies or beneficial to other researchers. For example, while a study of causal learning might focus on variables like experienced contingencies, delays, and player choice, another researcher might use player movement data to study player navigation or gaze direction to study attention. Even if an investigator does not foresee using these variables in pursuit of addressing their own research questions, making such data available in public data repositories enables other laboratories to address complementary research questions.

The ease with which behavioral data is collected in a game engine's virtual environment is in stark opposition to the technical and financial challenges within physical environments. Tracking a subject's location requires either (a) video recording of behavior followed by algorithmic determination of location using machine learning, (b) an apparatus that can track location based on floor pressure or break-beam detection, or (c) installation of a tracking system located on the subject's body (e.g., using RFID) and recorded by specialized hardware. Tracking individual behaviors (e.g., rearing, eating, jumping) is even more challenging. Recording gaze direction requires head mounts, eye tracking technology, or video recordings analyzed by machine learning. As a result of these challenges, these additional variables are rarely collected in most behavior analytic experiments.

The rich array of data available during video game play does present some challenges. The initial challenge is the need to visualize how values change over time. Secondly, some of these data are collected at very different time scales (e.g., every 100 ms for continuously changing variables like location vs. every few seconds or minutes for response choice), and these data would need to be integrated, if necessary, to address certain research questions. Furthermore, standard data analysis tools may be insufficient to assess temporal relationships. We note, however, that it is precisely those challenges that can drive innovation.

The collection of large datasets also opens the world of machine learning as a tool to reveal patterns in the data. Machine learning has had a limited impact in behavior analysis because the data sets are often too small – machine learning algorithms require large sets of data to avoid overfitting (i.e., developing a statistical description that does not generalize to new subjects or experimental variations). Recording data at a higher resolution overcomes this limitation.

Consequences of a Move to Video Game Engines

We would be remiss if we did not discuss the challenges inherent in adopting video game engines as a platform for understanding human behavior. A common concern is the need to learn to program. Indeed, programming skills are increasingly necessary for all experimental psychologists. Many experimental platforms – ranging from hardware-based systems like MED-PC used to drive Med-Associates operant chambers, to software solutions like PsychoPy, Psychtoolbox, and E-Prime – require familiarity with programming logic. Even when GUIs are available (e.g., PsychoPy's Builder), they inherently reflect scripting logic and may require inserting custom script components for advanced functionality. The same holds for data analyses using popular programming languages for statistical computing like R and Python. The question is thus not the need to program but rather the relative difficulty in doing so.

Video game engines are designed to make it easier to develop games by providing a large array of tools such that physics is respected (e.g., so that two objects cannot occupy the same space and that sounds dissipate with distance), that an avatar can only see objects within its field of view, and that rendering of objects is done efficiently in real-time. Furthermore, much of what goes into designing a game can be directly controlled through a engine's graphic user interface. Thus, many of the most complex programming tasks are already addressed.

Regardless, there is still a lot to learn if one chooses to program a game from the ground up. There are three ways to ease the task of game development for studying human behavior: a) modifying an existing game, b) downloading or buying assets from a game's asset store, and c) using the many online how-to resources for popular game engines. Now is an especially favorable time for adopting game engines in research. Modern engines like Unity and Unreal are widely used, well-documented, and supported by a wealth of tutorial videos, books, and example projects, and they are already employed by researchers across different fields. Of course, it is also beneficial to develop a collaboration with someone who developed a game that also suits one's needs. Game engines facilitate such collaborations by allowing researchers to work within the same platform, reducing the need to start from scratch or painstakingly explain technical details to collaborators.

For researchers with aspirations that go beyond their achievable skill levels, contractable game developers can also be hired to produce a customizable game. This approach requires grant funding or inspired administrators. To plan for the future, such a game should include mechanisms and interfaces that make it easy to modify the game's behavior. For example, a game used to study self-control could read a text file produced by the scientist that specifies the delays and magnitudes to be experienced at each decision location, whereas a game designed to study navigation could include an interface for the experimenter to change the locations of targeted rewards or the time intervals between data collections.

Once a scientist has tackled the programming challenge, the next task is to design the experiment (Rafferty et al., 2014). The free-flowing nature of gameplay unlocks a wide range of possible designs – unlike common tools like PsychoPy and E-Prime, which often rely on structured, trial-by-trial designs, game engines offer the flexibility to create continuous, open-ended 2D and 3D environments. A key advantage of using games is that they are often designed without a high degree of structure: a player can make consequential decisions while traveling throughout an environment, and stimuli can change continuously in time and space. This lack of strong constraints means that the scientist must design the task to avoid obvious confounds (e.g., a player always choosing the easiest task first) and to add restrictions where necessary (e.g., not allowing subjects to make a choice between targets from so far away that could not have discriminated the visual characteristics of the choices). Pilot testing becomes critical in identifying problematic behaviors, unproductive strategies, and unforeseen incentives that undermine the validity of the experiment. The selective application of restrictions, rewards, and negative consequences can increase the quality of the data obtained. Finally, it is also helpful to identify and eliminate game elements that are aversive to sustained game play (e.g., long travel times between game locations or excessively long periods of inactivity).

Once the data have been obtained, they are often more complex than is typical in a behavior analytic study (Allen et al., 2024). When starting to use video games, it may be easiest to store different types of behavior in different files (e.g., choices in one file, avatar location in another), but it is necessary to use consistent time stamps on data entries so that data files can be re-integrated, if necessary. This increased complexity will challenge both the application of current theories (which may not address continua or the behavioral features collected) and the analysis of the data. Traditional statistical techniques like the t-test and ANOVA are restricted to categorical predictors. Continuous or graded predictors require multilevel (aka mixed effects) modeling for within-subject designs, and unfamiliar techniques (e.g., spatial statistics or path similarity metrics) may be required to address research questions of interest.

These challenges should be considered spurs to innovate, not burdens to be avoided. Innovation typically unfolds incrementally, and no single experiment must tackle every issue outlined above. Only by anticipating these issues ahead of time and taking targeted steps to address them can researchers save later time and effort. The next section provides concrete examples of how video game engines have been used in our laboratory to study an array of behavioral questions.

Examples from Dr. Michael Young's laboratory

Classical conditioning using a causal inference paradigm

In the first game-based study to come out of Dr. Michael Young's laboratory, participants were tasked with assessing the predictive relationships between the actions of three non-player characters (NPCs) and an outcome of interest (an explosion, Young & Nguyen, 2009). The initial paper involved an outcome that could be delayed up to 2 s, the delays were either consistent or varied in duration, and the delays could be unfilled or filled with a bridging auditory stimulus. The subject needed to identify which NPC was using a weapon that caused explosions on the buildings that the subject was tasked to protect (Figure 1).

Figure 1. A snapshot of the targets ("orcs") in the causal choice task, one of which was firing a weapon creating the explosion on the tower. The green bar in the middle is the remaining life for the target in the player's crosshairs.

To increase the number of decisions that a subject must make, the NPCs were grouped in triplets at seven different locations within the game space (a hilly village with buildings to be protected). A subject was randomly dropped within the village and could move about the environment at will. Subjects could choose to watch the NPCs for as long as they wished and view from any perspective or distance that they chose – we made no attempt to control these factors since there was no reason to expect these factors to be confounded with the independent variables (IVs). Once a subject correctly identified the seven NPC targets causing damage, the game reset, the NPCs were respawned, but the assignment of IVs to target triplets refreshed so that subjects would have to rediscover the environmental contingencies for each NPC triplet.

Performance was evaluated by both accuracy and latency of each decision. Interestingly, in a later study (Young, Sutherland, Cole, et al., 2011), participants were able to leverage the speed accuracy tradeoff to avoid lower accuracy (e.g., when decisions were harder) by waiting longer to make their decision. Although avatar location was also recorded, these data were not formally analyzed; movement data did, however, allow the authors to address a reviewer's concern.

This custom video game was built from one of the two tutorial games provided with the Torque game engine that was used in producing the game. Most modifications of the tutorial were minor – adding more NPCs, replicating buildings, changing NPC and building locations, and reshaping the terrain by adding surrounding mountains to encourage subjects to stay in the village. The two largest changes to the tutorial game were to record key events in data files (firing of weapons and, every second, avatar location) and the manipulation of weapon characteristics (delay to explosion, etc.). Later studies included probabilistic outcomes, instituted time pressure, or explored the relationship between a subject's sex and performance (Nguyen et al., 2010; Young, 2014; Young & Cole, 2012; Young, Sutherland, & Cole, 2011).

Operant behavior in a delay discounting task

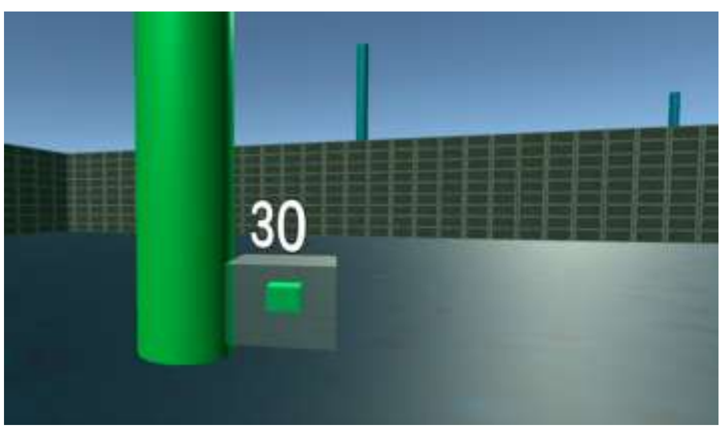
A second, even longer series of experiments involved the implementation of a type of delay discounting task in which players experienced the delays to the larger-later outcomes and the magnitude of the outcome had consequences (more damage to an NPC). A key modification in the basic design was to leverage the continuous nature of game dynamics by having a 'charge bar' that gradually filled (and thus the design was called an "escalating interest task") – thus, a player could fire rapidly when the weapon's bar had little charge and thus do little damage, or the player could wait and fire after the bar was at full charge and do much more damage (Figure 2).

Figure 2. A snapshot of the subject's perspective in another escalating interest task. In this example, there were two "charge bars" – one showed the amount of damage that a successful shot would inflict whereas the other showed the likelihood that a shot would work.

The initial research (Young, Webb, et al., 2011) was built on the tutorial game used in the laboratory's studies of causal inference. The player was tasked with destroying all of the NPCs in each game level, and each NPC required multiple shots to destroy thus creating a rich dataset for evaluation. A vast array of IVs were studied in these experiments including delay length, whether magnitude and/or probability increased with wait time, the temporal dynamics of the growth in magnitude, methods of training and their impact on wait times, and the impact of resource limitations (limited ammunition) on wait times (Webb & Young, 2015; Young & Hancock, 2025; Young & Howatt, 2022, 2023; Young & McCoy, 2015; Young et al., 2018; Young, Webb, Rung, et al., 2013; Young et al., 2014; Young, Webb, Sutherland, et al., 2013)

In later variations of the study of delay discounting, a second version of the game was developed by Brian Howatt from the ground up using the Unity game engine rather than starting with an existing game environment. The resulting game provided a framework from which other experiments in our lab could more easily be built. Creating the entire environment in Unity was also an excellent training platform for the student programming the game (Figure 3). Additionally, it was reassuring that the commonly observed behaviors in the first game were very similar to those observed in the new environment thus demonstrating generalization.

Figure 3. A snapshot of the subject's perspective in a later version of the escalating interest task. In this example, the target was an opaque cube that grew in size over time to a maximum size (the transparent enclosing cube). The box served the same informational purpose that the charge bar did in earlier versions of the escalating interest task.



Studying reinforcement schedules and variable magnitude using loot boxes

Inspired by the practice of loot boxes endemic to modern videogames, Patrick Hancock created a novel paradigm for studying the influence of schedule variability and variable reward magnitude on sustained operant responding. Loot box is an umbrella term for digital ‘mystery box’ that players can purchase to receive randomly-generated in-game rewards (Ballou et al., 2022; King & Delfabbro, 2018; Nielsen & Grabarczyk, 2018), with available rewards being idiosyncratic to each particular game. For older generations, opening a loot box is similar to opening a pack of baseball trading cards – you do not know what you are buying until the pack is opened. As such, loot boxes represent a modernized application of intermittent schedules of reinforcement that lead to persistent operant responding.

Using an open-source loot box simulator built in the Unity game engine by Kao (2019), Hancock modified the game to include a new focal task, track behavioral data, and most importantly, vary the timing of reinforcement according to different reinforcement schedules. In our first use of the game, we studied how operant responding (in this case, choosing to purchase loot boxes) changed as a function of the timing and predictability of reinforcement.

The focal task of the game was to earn points by clicking a red button at a rate of one point per click. After accruing enough points (in this example, 50 points), the game would offer participants the choice between using their points to open a loot box that could reward them with a much larger number of points (500) or to bank their points (Figure 4). Each time a participant earned 50 points by clicking, the option of opening a loot box was again presented. If participants opened the requisite number of loot boxes (as determined by the current variable ratio schedule of reinforcement), the box would open to reveal a stack of 500 gold coins (Figure 5). Otherwise, the box would contain a pile of rocks worth zero points. To beat the game, participants needed to earn a target number of points (5,000 points in our first study). We assumed that participants would strive to minimize completion time, thereby making the loot boxes a tempting chance at saving time. We also included an inventory panel that tracked each outcome, thus mitigating the potential confound of recall ability. In addition to our main outcome measure of participant choices, the game also recorded click locations, time of each click, and click durations for future exploratory analyses of response variability (i.e., variability in click location), interresponse times, and pre-ratio or post-reinforcement pauses (Felton & Lyon, 1966; Griffiths & Thompson, 1973).

While we initially developed the paradigm based on simple responses and manipulations (not unlike the traditional Skinner box), we have already begun to incorporate further complexity. In our second study, we added a small pile of bronze coins worth 50 points and a large pile of silver coins worth 100 points as outcomes in addition to the gold coins and rocks used in the first study. Although issues of variable reinforcer magnitude are prominent in the gambling literature (e.g., Delfabbro & Winefield, 1999; Dixon et al., 2010), magnitude manipulations are rare in behavioral analysis with human participants (with Berardi et al., 2020 representing the only contemporary example to our knowledge). Hancock's (2025) initial results revealed that variable reward magnitudes led to better participant engagement, thus increasing the effectiveness of this limited duration training.

Figure 4. The choice given to participants at the end of each trial in the loot box simulator. Choosing 'yes' will subtract 50 points from the participant's usable points and open the loot box (the chest). The chest contents are presented to the player and immediately added to the inventory on the right side of the screen; any earned points are added to the player's total.

Choosing 'no' leaves the chest closed, will subtract 50 points from the participant's usable points, and add fifty points to the player's banked points.

Figure 5. Example of a winning loot box. The 500 points were added to the banked points and, after the winning animation is complete, will appear as +500 in the right panel as a record of that win.

Foraging for resources in a virtual environment

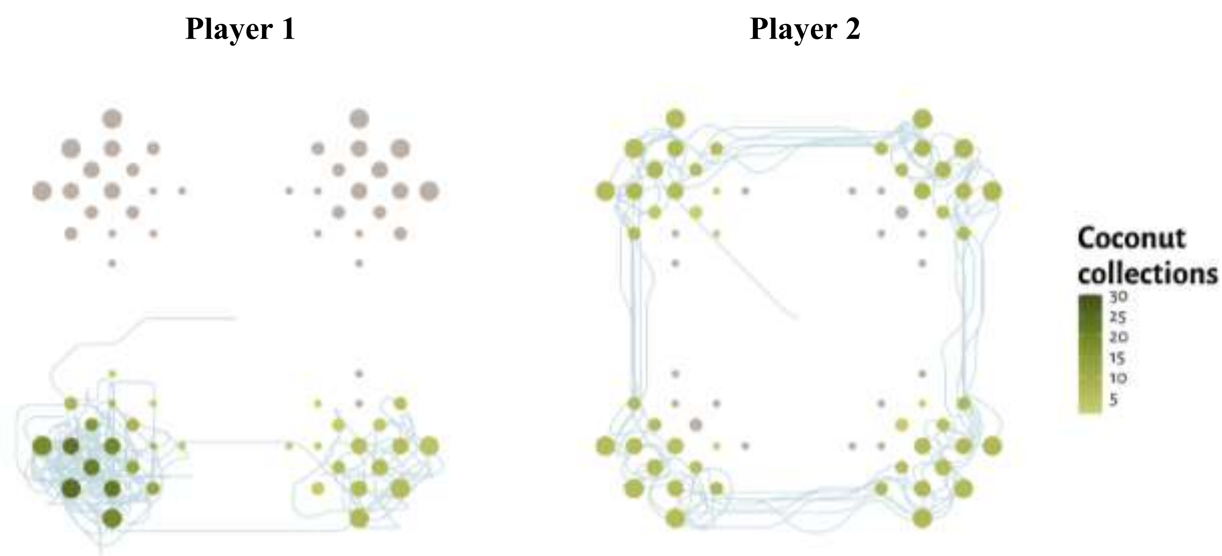
Video game engines also provide the capability to generate immersive, naturalistic environments. In another example from the lab, Luke Watson is using the Unity game engine to create a 3D, open-world foraging environment. Using a combination of code and community-created assets, he simulated the features of naturalistic foraging. Subjects navigated their avatar, a cartoon monkey, around a beach in search of coconuts distributed on the ground (Figure 6).

The subject earns points (which are necessary to complete each game level) by moving close enough to collect a coconut. Larger coconuts are worth more points. When collected, a coconut disappears and respawns in the same location after a fixed time interval thus modeling the process of resource replenishment as a function of time.

Figure 6. A subject's monkey avatar (center of the screenshot) collecting coconuts on the beach. Points accumulated ("Points 26") as well as a goal progress bar (location of a coconut on a gray horizontal bar) are displayed in the top left corner.

Our initial explorations using this task are investigating how variables like replenishment rate, spatial distribution, and coconut size affect foraging. Specifically, Luke is interested in how foraging patterns and their underlying cognitive strategies, operationalized as the routes taken to collect coconuts, develop over the course of the experiment and systematically vary as a function of our variables of interest. For instance, initial pilots have revealed interesting individual differences in routes taken and performance that vary in their efficiency (Figure 7; the participant on the right was slightly more efficient than the participant on the left, walking 30 fewer steps to collect the necessary points to complete the level). The use of a game engine allows us to capture high-fidelity movement data from our participants and enables the exploration of novel areas of human behavior – all at a low cost.

Figure 7. Example paths (light blue lines) taken by two pilot participants. A darker green coconut indicates that it was collected more times by the subject whereas a gray coconut indicates that it was never collected (see legend on right).



Multiplayer environments

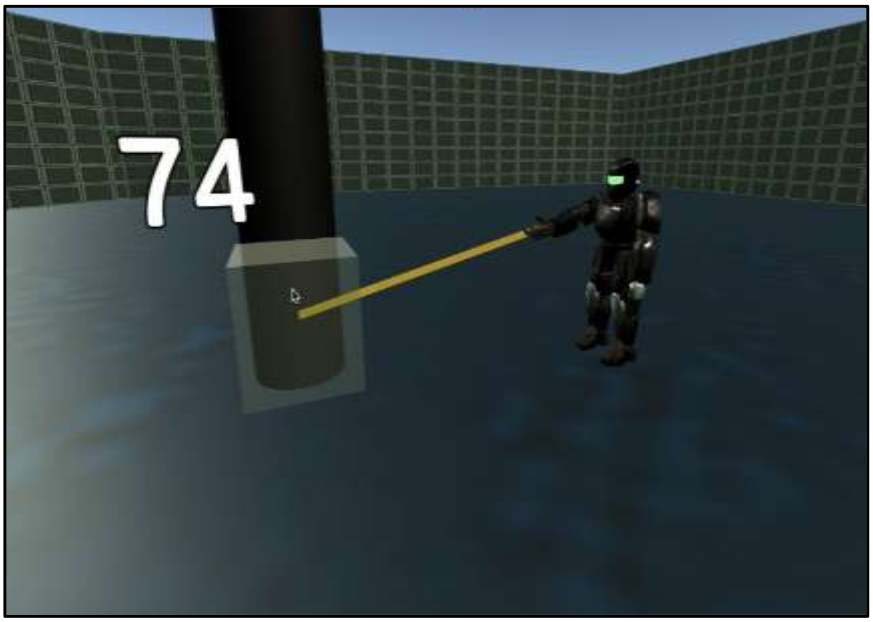
Video game engines can also be used to study mutually contingent reinforcement using both simulated and real co-players. Real-world reinforcement contingencies frequently involve multiple organisms each producing responses shaped by reinforcement, and their behavior will be sensitive to both their own reinforcement history and the anticipated actions of the other organisms. Including other participants into a task can help achieve the goals of accelerating behavior analysis research and creating richer environments with stronger generalizability. For example, problem behaviors in a classroom may be reinforced by the behavior of other students (e.g., praise or attention). Additionally, substance abusers may feel socially compelled to use drugs or even escalate their consumption to fit in with their peers.

Multiplayer experiments are not currently supported in experiment-building software like PsychoPy or E-Prime, but there is limited support in the Gorilla experiment builder (https://gorilla.sc). Researchers can host their own multiplayer experiments online, but this requires creating one's own web-based system using tools such as JavaScript and NodeJS that can quickly incur prohibitive costs and delays. In contrast, game engines can easily accommodate a wide variety of multiplayer designs regardless of their complexity. In contrast, game engines like Unity and Unreal offer robust support for multiplayer designs, including real-time interactions and extensive networking features, which can accommodate a wide variety of multiplayer designs directly without requiring web development

To study learning and behavior in multiplayer environments, we conducted a series of experiments in which we inserted computerized competitors into the delay discounting task (see Figure 8), and these competitors could steal rewards if participant waited too long to collect them (Howatt & Young, 2025). The competitors were robot avatars, and we manipulated their visibility, average response times, and predictability to study the degree to which these factors impacted individuals' willingness to wait for larger rewards. Despite being in a larger, dynamic, more open-world environment, we were still able to exercise strong experimental control over these computer avatars regarding their exact size, color, positioning, visibility, and response times to study how the relevant variables affected behavior. We maintained strong internal validity to test competing theories about the effects of the competitors while still enhancing the external validity of the results.

Although this specific example matched human participants with computerized competitors, most game engines are designed to extend tasks into multiplayer environments where human participants interact with each other using either a Local Area Network in a dedicated laboratory room or online where people can play from any location with internet access. The use of game engines thus opens exciting avenues of research that would be difficult to implement using more traditional methods.

Figure 8. Screenshot of the delay discounting task with a computerized competitor (the black-colored robot avatar). The yellow line signals that the competitor made a response to the target and obtained the points instead of the player.



Conclusion

> It is characteristic of the human species that successful action is automatically reinforced. The fascination of video games is adequate proof. What would industrialists not give to see their workers as absorbed in their work as young people in a video arcade? What would teachers not give to see their students applying themselves with the same eagerness? (Skinner, 1984, pp. 951-952)

Game designers are increasingly aware of the power of reinforcement principles and have leveraged them to increase game play. This extension of operant principles has not always been well-received since increasing the duration of game play is not the same as maximizing the game’s enjoyment (Søraker, 2016). This apparent conflict is the result of a particular operationalization of the goal of commercial gaming – increasing the amount of game play often in pursuit of monetization. However, operant principles can be applied to reinforce other behavioral repertoires once the hard work has been done to operationalize alternative goals like making a game more enjoyable, fun, challenging, or immersive. The more enjoyable the game, the more likely that someone will purchase it and continue to play.

Advancing our field’s understanding of reinforcement principles may be accelerated because video games not only allow but also encourage exploratory behaviors, and behavioral variation is a prerequisite for effective selection processes. Video game players can test more responses and outcomes in a game than they would in real life due to their extensive experience with the real world. For example, someone playing a game for the first time may test whether jumping off a building or walking through a deep pool of water causes damage to their avatar because some games implement these game mechanics whereas others do not. But, that same person would not evaluate the effects of jumping off a building or walking through a lake in the real world. This increased exploration may be a natural part of being in a new environment with untested assumptions or due to the lower cost of mistakes.

If one of the primary goals of the experimental analysis of behavior is to understand reinforcement principles and their generalizability, then gaming platforms can help the field make a major step forward. Behavioral change, however, can result from processes beyond reinforcement and punishment to include the learning of cue relations (e.g., compound cues), demand elasticity, behavioral variability, rule-governed behavior, and generalization and discrimination of stimuli, responses, and outcomes. Individually or collectively, the study of these phenomena using video game engines will enrich our understanding of human behavior.

References

Allen, K., Brändle, F., Botvinick, M., Fan, J. E., Gershman, S. J., Gopnik, A., Griffiths, T. L., Hartshorne, J. K., Hauser, T. U., & Ho, M. K. (2024). Using games to understand the mind. *Nature Human Behaviour*, *8*(6), 1035-1043.

Atak, J. R., & Nevin, J. A. (1983). Chromatic adaptation in the pigeon. *Animal Learning & Behavior*, *11*(4), 391-395.

Azoulay, P., Graff Zivin, J. S., Li, D., & Sampat, B. N. (2019). Public R&D investments and private-sector patenting: evidence from NIH funding rules. *The Review of Economic Studies*, *86*(1), 117-152.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *20*(4), 313-327. https://doi.org/10.1901/jaba.1987.20-313

Bailey, J. S. (1991). Marketing behavior analysis requires different talk. *Journal of Applied Behavior Analysis*, *24*(3), 445.

Ballou, N., Gbadamosi, C., & Zendle, D. (2022). The hidden intricacy of loot box design: A granular description of random mone-tized reward features. Proceedings of DiGRA 2022 Conference: Bringing Worlds Together,

Baron, A., & Galizio, M. (1976). Clock control of human performance on avoidance and fixed-interval schedules. *Journal of the Experimental Analysis of Behavior*, *26*(2), 165-180. https://doi.org/10.1901/jeab.1976.26-165

Berardi, V., Hovell, M., Hurley, J. C., Phillips, C. B., Bellettiere, J., Todd, M., & Adams, M. A. (2020). Variable magnitude and frequency financial reinforcement is effective at increasing adults' free-living physical activity. *Perspectives on Behavior Science*, *43*, 515-538.

Blackwell, A. A., Köppen, J. R., Whishaw, I. Q., & Wallace, D. G. (2018). String-pulling for food by the rat: Assessment of movement, topography and kinematics of a bilaterally skilled forelimb act. *Learning and Motivation*, *61*, 63-73.

Bouvier, P., Lavoué, E., & Sehaba, K. (2014). Defining engagement and characterizing engaged-behaviors in digital gaming. *Simulation & Gaming*, *45*(4-5), 491-507.

Brubaker, D. K., & Lauffenburger, D. A. (2020). Translating preclinical models to humans. *Science*, *367*(6479), 742-743. https://doi.org/10.1126/science.aay8086

Burgos-Robles, A., Kimchi, E. Y., Izadmehr, E. M., Porzenheim, M. J., Ramos-Guasp, W. A., Nieh, E. H., Felix-Ortiz, A. C., Namburi, P., Leppla, C. A., Presbrey, K. N., Anandalingam, K. K., Pagan-Rivera, P. A., Anahtar, M., Beyeler, A., & Tye, K. M. (2017). Amygdala inputs to prefrontal cortex guide behavior amid conflicting cues of reward and punishment. *Nature Neuroscience*, *20*(6), 824-835. https://doi.org/10.1038/nn.4553

Buskist, W., Morgan, D., & Terrell, D. J. (1985). On the experimental analysis of human behavior. *Mexican Journal of Behavior Analysis*, *11*(1 y 2), 69-78.

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, *54*(4), 297-312. https://doi.org/10.1037/h0040950

Case, D. A., Ploog, B. O., & Fantino, E. (1990). Observing behavior in a computer game. *Journal of the Experimental Analysis of Behavior*, *54*(3), 185-199.

Clayton, J. A., & Collins, F. S. (2014). Policy: NIH to balance sex in cell and animal studies. *Nature*, *509*(7500), 282-283. https://doi.org/10.1038/509282a

Critchfield, T. S., Doepke, K. J., Kimberly Epting, L., Becirevic, A., Reed, D. D., Fienup, D. M., Kremsreiter, J. L., & Ecott, C. L. (2017). Normative emotional responses to behavior analysis jargon or how not to use words to win friends and influence people. *Behavior Analysis in Practice*, *10*(2), 97-106. https://doi.org/10.1007/s40617-016-0161-9

Delfabbro, P. H., & Winefield, A. H. (1999). Poker-machine gambling: An analysis of within session characteristics. *British Journal of Psychology*, *90*(3), 425-439.

Dixon, M. J., Harrigan, K. A., Sandhu, R., Collins, K., & Fugelsang, J. A. (2010). Losses disguised as wins in modern multi-line video slot machines. *Addiction*, *105*(10), 1819-1824.

Dubova, M., Chandramouli, S., Gigerenzer, G., Grunwald, P., Holmes, W., Lombrozo, T., Marelli, M., Musslick, S., Nicenboim, B., Ross, L. N., Shiffrin, R., White, M., Wagenmakers, E. J., Burkner, P. C., & Sloman, S. J. (2025). Is Ockham's razor losing its edge? New perspectives on the principle of model parsimony. *Proceedings of the National Academy of Sciences*, *122*(5), e2401230121. https://doi.org/10.1073/pnas.2401230121

Entertainment Software Association. (2025). *Essential facts about the video game industry*. https://www.theesa.com/resources/essential-facts-about-the-us-video-game-industry/2025-data/

Epifânio, J. C., & da Silva, L. F. (2023). Embracing applied behavior analysis on a serious game design document model. *IEEE Access*, *11*, 72070-72087.

Fantino, E., & Case, D. A. (1983). Human observing: Maintained by stimuli correlated with reinforcement but not extinction. *Journal of the Experimental Analysis of Behavior*, *40*(2), 193-210.

Felton, M., & Lyon, D. O. (1966). The post-reinforcement pause. *Journal of the Experimental Analysis of Behavior*, *9*(2), 131-134.

Flynn, M. K., & Rocheleau, C. (2024). Undergraduate research pools in psychology departments: Prevalence and practices. *Teaching of Psychology*, *51*(1), 39-45.

Freedman, D. H. (2016). Improving public perception of behavior analysis. *The Behavior Analyst*, *39*(1), 89-95.

Galizio, M. (2020). JEAB: Past, present, and future. *Journal of the Experimental Analysis of Behavior*, *113*(1), 3-7. https://doi.org/10.1002/jeab.574

Greem, D. M., & Luce, R. D. (1974). Variability of magnitude estimates: A timing theory analysis. *Perception & Psychophysics*, *15*(2), 291-300.

Griffiths, M. (1993). Fruit machine gambling: The importance of structural characteristics. *Journal of Gambling Studies*, *9*(2), 101-120.

Griffiths, R. R., & Thompson, T. (1973). The post-reinforcement pause: A misnomer. *The Psychological Record*, *23*, 229-235.

Hancock, P. M. (2025). *Developing an experimental paradigm for the study of loot boxes* [Master's thesis, Kansas State University].

Harrison, B., & Roberts, D. L. (2015). An analytic and psychometric evaluation of dynamic game adaption for increasing session-level retention in casual games. *IEEE Transactions on Computational Intelligence and AI in Games*, *7*(3), 207-219.

Harzem, P. (1984). Experimental analysis of individual differences and personality. *Journal of the Experimental Analyis of Behavior*, *42*(3), 385-395. https://doi.org/10.1901/jeab.1984.42-385

Henderson, V. C., Kimmelman, J., Fergusson, D., Grimshaw, J. M., & Hackam, D. G. (2013). Threats to validity in the design and conduct of preclinical efficacy studies: A systematic review of guidelines for in vivo animal experiments. *PLoS Med*, *10*(7), e1001489. https://doi.org/10.1371/journal.pmed.1001489

Hendrix, M., Bellamy-Wood, T., McKay, S., Bloom, V., & Dunwell, I. (2018). Implementing adaptive game difficulty balancing in serious games. *IEEE Transactions on Games*, *11*(4), 320-327.

Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The 'real-world approach' and its problems: A critique of the term ecological validity. *Frontiers in Psychology*, *11*, 721. https://doi.org/10.3389/fpsyg.2020.00721

Howatt, B. C., & Young, M. E. (2025). How competition shapes preferences for smaller sooner rewards. Available at SSRN: https://ssrn.com/abstract=5269986

James, R. J., O'Malley, C., & Tunney, R. J. (2016). Why are some games more addictive than others: The effects of timing and payoff on perseverance in a slot machine game. *Frontiers in Psychology*, *7*, 46.

Jarmolowicz, D. P., Reed, D. D., Stancato, S. S., Lemley, S. M., Sofis, M. J., Fox, A., & Martin, L. E. (2020). On the discounting of cannabis and money: Sensitivity to magnitude vs. delay. *Drug and Alcohol Dependence*, *212*, 107996. https://doi.org/10.1016/j.drugalcdep.2020.107996

Juul, J. (2012). *A casual revolution: Reinventing video games and their players*. MIT press.

Kao, D. (2019). Infinite loot box: A platform for simulating video game loot boxes. *IEEE Transactions on Games*, *12*(2), 219-224.

King, D. L., & Delfabbro, P. H. (2018). Predatory monetization schemes in video games (eg 'loot boxes') and internet gaming disorder.

Klapes, B., Calvin, O. L., & McDowell, J. J. (2020). A discriminated rapid-acquisition laboratory procedure for human continuous choice. *Journal of the Experimental Analysis of Behavior*, *114*(1), 142-159. https://doi.org/10.1002/jeab.612

Lin, H., Werner, K. M., & Inzlicht, M. (2021). Promises and Perils of Experimentation: The Mutual-Internal-Validity Problem. *Perspectives on Psychological Science*, *16*(4), 854-863. https://doi.org/10.1177/1745691620974773

Macaskill, A. C., Thompson-Davies, K., & Hunt, M. J. (2023). Reduced access to reinforcement drives delay discounting during experienced delays. *Journal of the Experimental Analysis of Behavior*, *120*(3), 416-428. https://doi.org/10.1002/jeab.883

Madden, G. J., & Perone, M. (2003). Effects of alternative reinforcement on human behavior: the source does matter. *Journal of the Experimental Analysis of Behavior*, *79*(2), 193-206. https://doi.org/10.1901/jeab.2003.79-193

Maes, E., De Filippo, G., Inkster, A. B., Lea, S. E., De Houwer, J., D'Hooge, R., Beckers, T., & Wills, A. J. (2015). Feature- versus rule-based generalization in rats, pigeons and humans. *Animal cognition*, *18*(6), 1267-1284. https://doi.org/10.1007/s10071-015-0895-8

Mills, W. (1899). The nature of animal intelligence and the methods of investigating it. *Psychological Review*, *6*, 262-274. https://doi.org/https://doi.org/10.1037/h0074808

Morford, Z. H., Witts, B. N., Killingsworth, K. J., & Alavosius, M. P. (2014). Gamification: The intersection between behavior analysis and game design technologies. *Behavior Analyst*, *37*(1), 25-40. https://doi.org/10.1007/s40614-014-0006-1

Nelson, R. R. (1959). The simple economics of basic scientific research. *Journal of Political Economy*, *67*(3), 297-306.

Newall, P., Hayes, T., Weiss-Cohen, L., James, R. J., Byrne, C. A., & Muggleton, N. (2025). Could the provision of multiple game themes be a neglected gambling product structural characteristic? Results of an online simulated gambling task. *International Gambling Studies*, 1-20.

Nguyen, N., Young, M. E., & Cole, J. J. (2010). The effect of number of options on choices involving delayed causation. *American Journal of Psychology*, *123*(4), 455-465.

Nielsen, R. K. L., & Grabarczyk, P. (2018). Are loot boxes gambling? Random reward mechanisms in video games. Proceedings of DiGRA 2018 Conference: The Game is the Message,

Pavitt, K. (1991). What makes basic research economically useful? *Research Policy*, *20*(2), 109-119.

Pedersen, M. K., Díaz, C. M. C., Wang, Q. J., Alba-Marrugo, M. A., Amidi, A., Basaiawmoit, R. V., Bergenholtz, C., Christiansen, M. H., Gajdacz, M., & Hertwig, R. (2023). Measuring cognitive abilities in the wild: Validating a population-scale game-based cognitive assessment. *Cognitive Science*, *47*(6), e13308.

Pineño, O. (2014). ArduiPod Box: A low-cost and open-source Skinner box using an iPod Touch and an Arduino microcontroller. *Behavior Research Methods*, *46*, 196-205.

Pound, P., & Ritskes-Hoitinga, M. (2018). Is it possible to overcome issues of external validity in preclinical animal research? Why most animal models are bound to fail. *Journal of Translational Medicine*, *16*, 1-8.

Rafferty, A. N., Zaharia, M., & Griffiths, T. L. (2014). Optimally designing games for behavioural research. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *470*(2167), 20130828.

Rafner, J., Biskjær, M. M., Zana, B., Langsford, S., Bergenholtz, C., Rahimi, S., Carugati, A., Noy, L., & Sherson, J. (2022). Digital games for creativity assessment: Strengths, weaknesses and opportunities. *Creativity Research Journal*, *34*(1), 28-54.

Ramnerö, J., Molander, O., Lindner, P., & Carlbring, P. (2019). What can be learned about gambling from a learning perspective? A narrative review. *Nordic Psychology*, *71*(4), 303-322.

Reed, P., Smale, D., Owens, D., & Freegard, G. (2018). Human performance on random interval schedules. *Journal of Experimental Psychology: Animal Learning and Cognition*, *44*(3), 309.

Richter, S. H., Garner, J. P., Auer, C., Kunert, J., & Würbel, H. (2010). Systematic variation improves reproducibility of animal experiments. *Nature Methods*, *7*, 167-168.

Richter, S. H., Garner, J. P., & Würbel, H. (2009). Environmental standardization: Cure or cause of poor reproducibility in animal experiments? *Nature Methods*, *6*, 257-261.

Saville, B. K., Epting, L. K., & Buskist, W. (2002). Selected publication trends in JEAB: Implications for the vitality of the experimental analysis of behavior. *Behavior Analyst*, *25*(1), 45-55. https://doi.org/10.1007/BF03392044

Shimizu, T. (1998). Conspecific recognition in pigeons (Columba livia) using dynamic video images. *Behaviour*, *135*(1), 43-53.

Skinner, B. F. (1930). On the conditions of elicitation of certain eating reflexes. *Proceedings of the National Academy of Sciences*, *16*(6), 433-438.

Skinner, B. F. (1957). The experimental analysis of behavior. *American Scientist*, *45*(4), 343-371.

Skinner, B. F. (1984). The shame of American education. *American Psychologist*, *39*(9), 947.

Smith, J. M. (2016). Strategies to position behavior analysis as the contemporary science of what works in behavior change. *The Behavior Analyst*, *39*, 75-87.

Søraker, J. H. (2016). Gaming the gamer? The ethics of exploiting psychological research in video games. *Journal of Information, Communication and Ethics in Society*, *14*(2), 106-123.

Steele, C. C., Gwinner, M., Smith, T., Young, M. E., & Kirkpatrick, K. (2019). Experience Matters: The Effects of Hypothetical versus Experiential Delays and Magnitudes on Impulsive Choice in Delay Discounting Tasks. *Brain Sci*, *9*(12). https://doi.org/10.3390/brainsci9120379

Stevens, S. S. (1971). Issues in psychophysical measurement. *Psychological Review*, *78*(5), 426.

Timberlake, W. (1988). The behavior of organisms: Purposive behavior as a type of reflex. *Journal of the Experimental Analysis of Behavior*, *50*(2), 305.

Toneatto, T., Blitz-Miller, T., Calderwood, K., Dragonetti, R., & Tsanos, A. (1997). Cognitive distortions in heavy gambling. *Journal of Gambling Studies*, *13*, 253-266.

Ushitani, T., & Fujita, K. (2005). Pigeons do not perceptually complete partly occluded photos of food: an ecological approach to the “pigeon problem”. *Behavioural Processes*, *69*(1), 67-78.

Vorderer, P., Hartmann, T., & Klimmt, C. (2003). Explaining the enjoyment of playing video games: the role of competition. Proceedings of the second international conference on Entertainment computing,

Wasserman, E., Kiedinger, R., & Bhatt, R. (1988). Conceptual behavior in pigeons: Categories, subcategories, and pseudocategories. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*(3), 235.

Webb, T. L., & Young, M. E. (2015). Waiting when both certainty and magnitude are increasing: Certainty overshadows magnitude. *Journal of Behavioral Decision Making*, *28*, 294-307. https://doi.org/10.1002/bdm.1850

Wittek, N., Wittek, K., Keibel, C., & Güntürkün, O. (2023). Supervised machine learning aided behavior classification in pigeons. *Behavior Research Methods*, *55*(4), 1624-1640.

Yee, N. (2006). Motivations for play in online games. *CyberPsychology & Behavior*, *9*(6), 772-775.

Yee, N. (2014). *The Proteus paradox: How online games and virtual worlds change us-and how they don't*. Yale University Press.

Young, M. E. (2014). Sex differences in the inference and perception of causal relations within a video game. *Frontiers in Psychology*, *5*, 926.

Young, M. E., Beckmann, J. S., & Wasserman, E. A. (2006). Pigeons' discrimination of Michotte's launching effect. *Journal of the Experimental Analysis of Behavior*, *86*(2), 223-237. https://doi.org/10.1901/jeab.2006.60-05

Young, M. E., & Cole, J. J. (2012). Human Sensitivity to The Magnitude and Probability of a Continuous Causal Relation in a Video Game. *Journal of Experimental Psychology-Animal Behavior Processes*, *38*(1), 11-22. https://doi.org/10.1037/a0026357

Young, M. E., & Hancock, P. M. (2025). Maximizing within-session stability in individual differences during an experiential impulsivity task. *Learning & Behavior*. https://doi.org/10.3758/s13420-025-00677-0

Young, M. E., & Howatt, B. C. (2022). Learning when to wait and when to act. *Learning and Behavior*, *50*(2), 207-221. https://doi.org/10.3758/s13420-021-00482-5

Young, M. E., & Howatt, B. C. (2023). Cue transparency leads to a myopic decision habit: Optimal training ratios to avoid overshadowing. *Learning and Motivation*, *82*. https://doi.org/https://doi.org/10.1016/j.lmot.2023.101886

Young, M. E., & McCoy, A. W. (2015). A delay discounting task produces a greater likelihood of waiting than a deferred gratification task. *Journal of the Experimental Analysis of Behavior*, *103*, 180-195. https://doi.org/10.1002/jeab.119

Young, M. E., & Nguyen, N. (2009). The problem of delayed causation in a video game: Constant, varied, and filled delays. *Learning and Motivation*, *40*, 298-312.

Young, M. E., Sutherland, S. C., Cole, J., & Nguyen, N. (2011). Waiting to decide helps in the face of whether uncertainty but not when uncertainty. *Learning & Behavior*, *39*, 115-124.

Young, M. E., Sutherland, S. C., & Cole, J. J. (2011). Individual differences in causal judgment under time pressure: Sex and prior video game experience as predictors. *International Journal of Comparative Psychology*, *24*, 76-98.

Young, M. E., Vangsness, L., & McCoy, A. W. (2018). The temporal dynamics of waiting when reward is increasing. *Behavioural Processes*, *149*, 16-26. https://doi.org/10.1016/j.beproc.2018.01.015

Young, M. E., Webb, T. L., & Jacobs, E. A. (2011). Deciding when to “cash in” when outcomes are continuously improving: An escalating interest task. *Behavioural Processes*, *88*, 101-110.

Young, M. E., Webb, T. L., Rung, J. M., & Jacobs, E. A. (2013). Sensitivity to changing contingencies in an impulsivity task. *Journal of the Experimental Analysis of Behavior*, *99*, 335-345.

Young, M. E., Webb, T. L., Rung, J. M., & McCoy, A. W. (2014). Outcome probability versus magnitude: When waiting benefits one at the cost of the other. *PLOS One*, *9*(6), e98996. https://doi.org/10.1371/journal.pone.0098996

Young, M. E., Webb, T. L., Sutherland, S. C., & Jacobs, E. A. (2013). Magnitude effects for experienced rewards at short delays in the escalating interest task. *Psychonomic Bulletin & Review*, *20*(2), 302-309. https://doi.org/10.3758/s13423-012-0350-7

Zeigler, H. P., & Feldstein, R. (1971). A feedometer for the pigeon. *Journal of the Experimental Analysis of Behavior*, *16*(2), 181-187. https://doi.org/10.1901/jeab.1971.16-181