

Article

Addictive Games: Case Study on Multi-Armed Bandit Game

Xiaohan Kang ^{1,*} , Hong Ri ¹ , Mohd Nor Akmal Khalid ^{1,2} , Hiroyuki Iida ^{1,2} 

¹ School of Information Science, Japan Advanced Institute of Science and Technology, Nomi 923-1292, Japan; s1910192@jaist.ac.jp (H.R.); akmal@jaist.ac.jp (M.N.A.K.); iida@jaist.ac.jp (H.I.)

² Research Center for Entertainment Science, Japan Advanced Institute of Science and Technology, Nomi 923-1292, Japan

* Correspondence: kangxiaohan@jaist.ac.jp

Abstract: The attraction of games comes from the player being able to have fun in games. Gambling games that are based on the Variable-Ratio schedule in Skinner's experiment are the most typical addictive games. It is necessary to clarify the reason why typical gambling games are simple but addictive. Also, the Multiarmed Bandit game is a typical test for Skinner Box design and is most popular in the gambling house, which is a good example to analyze. This article mainly focuses on expanding on the idea of the motion in mind model in the scene of Multiarmed Bandit games, quantifying the player's psychological inclination by simulation experimental data. By relating with the quantification of player satisfaction and play comfort, the expectation's feeling is discussed from the energy perspective. Two different energies are proposed: player-side (E_r) and game-side energy (E_i). This provides the difference of player-side (E_r) and game-side energy (E_i), denoted as E_d to show the player's psychological gap. Ten settings of mass bandit were simulated. It was found that the setting of the best player confidence (E_r) and entry difficulty (E_i) can balance player expectation. The simulation results show that when $m = 0.3, 0.7$, the player has the biggest psychological gap, which expresses that player will be motivated by not being reconciled. Moreover, addiction is likely to occur when $m \in [0.5, 0.7]$. Such an approach can also help the developers and educators increase edutainment games' efficiency and make the game more attractive.



Citation: Kang, X.; Ri, H.; Khalid, M.N.A.; Iida, H. Addictive Games: Case Study on Multi-Armed Bandit Game. *Information* **2021**, *12*, 521. <https://doi.org/10.3390/info12120521>

Academic Editor: Gabriele Gianini

Received: 21 November 2021

Accepted: 11 December 2021

Published: 15 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: game-based learning; player satisfaction model; game experience; game refinement theory; multiarmed bandit game

1. Introduction

In the development of games, player motivation is always the goal object for game designers. Reward motivation can stimulate the pursuit to achieve the goals that are often used for behavior guidance in many areas such as business, education, human resource management, to name a few. A representative work on behaviorism by Skinner believed that after a specific behavior is rewarded, the specific behavior will be strengthened and solidified after continuous reinforcement [1]. The rules, conditions, and intensity of reward will also affect the incentive mechanism's effectiveness and the driving force of behavior. In previous clinical studies conducted on the animal, the dopamine system in the brain is associated with Beta signal, which is related to the activation of the orbitofrontal cortex when confronted with rewarding activities such as getting food [2]. Interestingly, similar results were obtained in human experiments [3]. However, different from animals, humans are good at learning how to predict the recurrence of reward signals [4,5].

Gambling games that typically have the highest uncertainty in games are typical reward-driven games. Usually, the result cannot be determined before placing a bet, and the game starts after stopping the betting. So the reward mechanism in gambling games expresses an immense appeal on players, and there is a definite possibility to cause addiction in the player. The mechanism of gambling games relies on the reward mechanism, and an instant reward feedback mechanism makes the player have a great curiosity to win the game.

Moreover, the reward setting of gambling based on reinforcement schedules makes the game more unpredictable and heavily reliant on information uncertainty. In terms of physiological mechanisms [6], rewards lead to the secretion of dopamine, which gives players a sense of pleasure. Also, the body's physiological mechanisms always seek for dopamine release repeatedly, at any cost, making it keen to explore and try new things, and this has an escalating effect on the player's motivation. The game has clear and specific goals, and each time a player completes a challenge, he or she is rewarded with a reward that disappears in the form of obstacles such as enemies, increases in experience and ability, an extension of the challenge time, or the opening of the next level. This situation is immediate, continuous and varied, and has an essential motivational impact on the player [7].

The first Multiarmed Bandit game is the mechanical slot machine called the Liberty Bell with three spinning reels, which was invented in 1895 by a car mechanic, Charles Fey [8] which then became one of the most popular slot machines in the gambling house. Such phenomenon acts as the motivation to adopt Multiarmed Bandit games for conducting analysis relative to the player's perceived psychology of the rewards obtained from such game. Game refinement (GR) theory, which was first introduced by Iida et al. [9], proposes the idea of analyzing and understanding game progress based on the uncertainty of the game result. It is a crucial evaluation standard and plays an essential role in every different game field. Based on variable ratio schedules, the player satisfaction model [10] provides a link between game refinement theory and reinforcement schedules. By connecting with the reward ratio (say N), the game's energy could be calculated, which shows how much the game satisfied the player based on the reward mechanism.

However, previous studies mainly focused on classifying players based on their motivations. For example, Malone initially believed that entertainment motivations are divided into three categories: challenge, fantasy, and curiosity [11]. The three types of motivations complement each other and are the deep reasons why humans like games. In Bartle's player model, motivations are analyzed to classify players, but the model does not explain the motivations of multiplayer games itself [12].

Meanwhile, the theory of motivation in game-playing makes up for the shortcomings of the MUD player model and analyzes multiplayer through five motives as the classification factors [13]. However, these models are based on classification, and there is no motivation analysis based on quantification. The quantitative psychological gap proposed in this paper analyzes the difference between players' expectations and reality by computational methods. Thus, the classification of players' intrinsic motivation comes from their confidence or unwillingness towards a reward.

The objective of this paper is two-fold: firstly, the reward mechanism of the gambling machine, the Multiarmed Bandit games, is presented for the first time to clarify it via the player satisfaction model. Secondly, such a model attempts to quantify the player's psychology during the game and analyze the underlying reasons for addiction. The main research questions for the paper is why the gambling game addictive and is it able to estimate players' psychological inclination?

Therefore, to the best of our knowledge, the player's psychological gap changes are defined for the first time based on the motion in mind model. Furthermore, analysis via simulation data of the two Multiarmed Bandit games under different settings was conducted to determine the player's psychological tendency. The experiment verifies the computational method of the player's psychological tendency where its potential applications were outlined.

2. Theoretical Framework

2.1. Multiarmed Bandit

A Multiarmed Bandit [14] is an example of a classical game for the gambler's psychology. Balancing the benefits of exploration and exploitation demonstrates the impact of uncertainty on future decisions. A gambler is presented with several slot machines

without knowing to advance each slot machine's actual profit. Each device provides a random reward from a probability distribution specific to that machine. The gambler aims to maximize the sum of rewards earned through a sequence of lever pulls [15,16]. In this situation, based on the actual representation of the prior and posterior probabilities, the gambler will have an expected reward before each choice act is performed and will receive feedback, i.e., a real bonus, after completing the action. There will be a difference between the two rewards causing uncertainty in this game, which affects the judgment of performing the next choice and continuing the action. An appropriate psychological difference will stimulate the player's behavior, while too little or too much will reduce the player's interest in the game and affect the player's game life.

During the game process of the Multiarmed Bandit game, the crucial trade-off the gambler faces at each trial is between "exploitation" of the machine with the highest expected payoff and "exploration" to get more information about the anticipated profits of the other devices. The expectation and variance of winning money in each slot machine are different. The player would need to choose the slot every time to maximize the revenue.

An example of reward distribution for the 10-arm bandit is shown in Figure 1, where reward each time was obtained from the sampling results of the Gaussian distributions [17]. Each of the violin plots corresponds to a different Gaussian distribution, with their respective mean and variance values. The actual probability would be the winning probability of such a mechanism. The action values $q_*(a)$, $a = 1, \dots, 10$ were chosen according to a normal (Gaussian) distribution with mean 0 and variance 1 of the normal (Gaussian) distribution to be chosen.

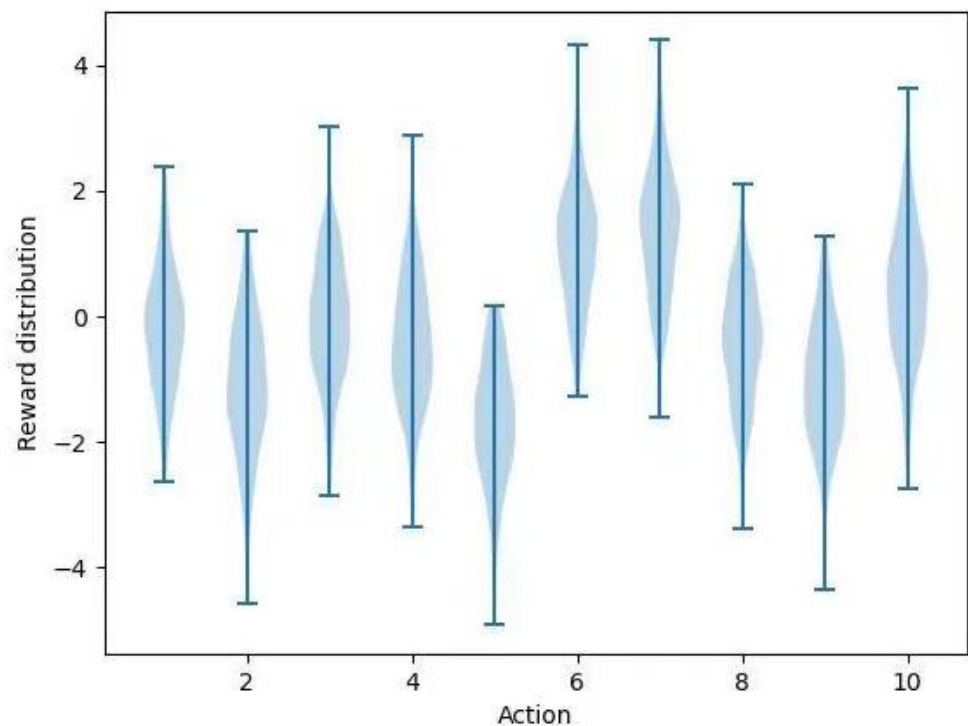


Figure 1. An example bandit from the 10 arms tested. Gray distributions express distribution with mean zero and unit variance. True value of each of 10 actions was selected among 10 choices (adopted from [18]).

2.2. Reward Mechanism in Games

Reinforcers are stimuli that could select appropriate behaviors and teach the player what to do [1]. The reward is one of the positive reinforcers [19]. As an essential feature of games, rewards exist in all types of games. Rewards come in many shapes and sizes, and if done right, can significantly increase the enjoyment and longevity of the game. Skinner's experiments on operant conditioning revealed reward on behavior reproduction, known as reinforcement theory. The reward schedule leads to the enjoyment of the game itself.

Mcgonigal [20] quoted: "The reward mechanism can help us to improve through random obstacles linked to our performance and better feedback mechanisms to make us work harder". It was used in many areas such as business [21], managements [22], educational areas [7], and so on. Specifically, psychological needs such as satisfaction may be associated with various feedback mechanisms provided by a game to the player. However, most of them focus on the reward mechanism itself, while few studies focus on the reward acquisition's uncertainty. As reward causes encouragement, uncertainty of a reward makes a situation thrilling, sense of crisis or urgency, and stimulate motivation [23].

Fiorillo et al. [24] examined the influence of reward probability and uncertainty on the activity of primate dopamine neurons. They found that the effect was greatest when reward uncertainty was 50 percent. Human studies on fMRI also reported evidence for a similar relationship between reward and uncertainty [25]. In addition, studies showed that large amounts of dopamine are released in uncertain situations of long-term uncertainty and significant rewards. This increase in dopamine output may contribute to the rewarding properties of gambling, with increased dopamine release during gaming and gambling-like tasks [25]. These studies suggested that reward uncertainty is indeed the key to player interest by controlling the uncertainty of reward and observing the dopamine levels and other neural signals.

On this basis, this paper intends to study further how the uncertainty of reward affects players' interest and leads to addiction at the psychological level of players. The Multi-armed Bandit game is based on a variable ratio schedule. Based on previous work [10,26], the game speed of the Multiarmed Bandit game is $1/N$ (N is the average ratio of the reward), which means that the average of N times attempts in the game would reinforce the player. The risk frequency ratio m , which is the risk frequency over the whole game length is defined as $m = 1 - 1/N = (N - 1)/N$. As such, this study explores the players' entertainment effect by analyzing the reward frequency (which will be discussed in detail in the next section).

2.3. Motions in Mind and Internal Energy Change in Games

Games are earning processes where players learn and adapt to grasp the rules of the game. Similarly, reinforcement schedules, which were explored by Skinner [27], were widely used in the learning environment. Based on such circumstances, game settings become essential factors that affect the player's experience [26]. Analogical links between motions in physics and motions in mind had been previously established based on the notions of winning rate (or velocity) v and winning hardness m , where the correspondence between the physics model and the game progress models is established based on the assumption of zero-sum game setting (Table 1).

Table 1. Analogical link between game and physics [26].

Notation	Game Context	Notation	Physics Context
y	solved uncertainty	x	displacement
t	progress or length	t	time
p	win rate	v	velocity
m	win hardness	M	mass
a	acceleration	g	gravitational acceleration

According to the game progress model, the slope (v) of $y(t) = vt$ of a game progress model has a contradictory relationship to m . In the current context, v is generally implying the rate of solving uncertainty, whereas m implies the difficulty of solving such uncertainty ($m = 1 - v$) [26]. Such correspondence enables indication of “physics in mind” in various games, specifically on three quantities: potential energy, momentum, and force. The potential energy (E_p) in the game is defined as the game playing potential or the expected game information required to finish a play [26], given by (1). At the same time, m is the game ‘mass’ (associated with the difficulty of solving the uncertainty), and v is the ‘velocity’ (associated with the rate of solving the uncertainty). According to the potential energy, player satisfaction could be expressed by employing reward mechanisms, and the “gravity” implied on such mechanism to the player [10], while $v = 1/N$.

$$E_p = 2mv^2 \tag{1}$$

Definition 1. Internal Energy Change (ΔU) in real-world physics can be defined as [28]: “For a closed system, with matter transfer excluded, the changes in internal energy are due to heat transfer (Q) and due to thermodynamic work (W) done by the system on its surroundings”. Accordingly, the internal energy change (ΔU) for a process is written as (2).

$$\Delta U = Q - W \text{ (closed system, no transfer of matter)} \tag{2}$$

Internal Energy Change definition provides the basis for this paper, which explores the formulation of internal energy change concerning the games. To define the change in Internal Energy Change of a game, we first need to clarify the concept of internal energy in relation to games. In this paper, we assume that the play process is metaphorically a closed system composed of the game and the player, where the heat transfer (Q) is the player-side energy associated with the expectation from the player. In contrast, the thermodynamic work (W) is associated with the game’s feedback (or game-side energy). Relative to the motion in mind model, the internal energy related to the changes in energy difference will be discussed further in the subsequent section.

3. Methodology

3.1. Energy Difference in Games

Two distinct energies were considered with a focus on player-side actual probability and game-side intuitive probability. The player-side energy E_i focused on the mass and velocity with a value of intuitive probability, whereas the game-side energy E_r based on the mass and velocity with a value of return rate. E_i and E_r are given in (3) and (4), where v_i and v_r stands for the intuitive probability and return rate respectively, hence $m_i + v_i = 1$ and $m_r + v_r = 1$ hold.

$$E_i = 2m_i v_i^2 \tag{3}$$

$$E_r = 2m_r v_r^2 \tag{4}$$

Table 2 provides the comparison of the two potential energies. The energy difference E_d is given by (5), which shows the player psychological discrepancy caused by the velocity difference between player and game.

$$E_d = E_i - E_r \tag{5}$$

Table 2. Two potential energies compared.

Notation	Game-Side	Player-Side
E_i	intuitive probability based game velocity	entry difficulty
E_r	return rate based game velocity	engagement(confident)

Remark 1. When $E_d > 0$, player confidence influences more profoundly than the game side, reflecting player confidence in gambling games. When $E_d < 0$, the game side influences more profoundly than the player side, reflecting high entry difficulty and causing player frustration similar to the one experiencing in chess-like games.

This study assumes that player-side energy (E_i) is based on the intuitive probability, which is the prior probability before every choice. Correspondingly, game-side energy (E_r) is based on the return rate, associated with the actual probability after a choice was made.

3.2. Upper Confidence Bound Method

UCB (Upper Confidence Bound) is a method first proposed by Lai and Robbins [29] that utilizes upper confidence values for dealing with the exploration-exploitation dilemma in the Multiarmed Bandit problem. The gambler's goal is to win more money and get the greatest return.

The algorithm steps: first try it for each arm, then at any moment calculate the score for each arm according to the following Formula (7), and select the arm with the largest score as the choice. Next, observe the selection results and update t and $n_{i,t}$, where $\hat{\mu}_{i,t}$ denotes the average reward obtained from the slot machine i with $i_t \in [1, 2, \dots, N]$, followed by $\sqrt{\frac{\ln t}{n_{i,t}}}$ being called the bonus, which is essentially the standard deviation of the mean, is the number of trials so far, and t is the number of times i was played.

Upper confidence bound (UCB) algorithms provide a simple but efficient heuristic approach to bandit problems [30]. In this study, UCB method was employed to simulate the player selection process. The predicted reward and actual reward of every step are counted during 10,000 times training in the experiments. At each round, the UCB algorithm would select the arm with the highest empirical reward estimate up to that point plus some term that is inversely proportional to the number of times the arm was played.

More formally, define $n_{i,t}$ as the number of times arm i was played up to time t . Then, $r_t \in [0, 1]$ denotes the reward observed at time t , while $i_t \in [1, 2, \dots, N]$ is the choice of the arm at time t . Then, the empirical reward estimate of arms i at time t is shown in (6). UCB assigns the following value to each arm i at each time t as shown in (7).

$$\hat{\mu}_{i,t} \in \frac{\sum_{s=1: I_s=i}^t r_s}{n_{i,t}} \quad (6)$$

$$UCB_{i,t} := \hat{\mu}_{i,t} + \sqrt{\frac{\ln t}{n_{i,t}}} \quad (7)$$

To briefly describe the UCB Algorithm 1, the following were the steps involved:

- Initialize the number of round, random generator, and arm choices (line 12–17). Then, try it for each arm (line 18).
- Calculate the score for each arm randomly (line 13–15) and according to Formula (7) (line 20–21), of which the arm with the largest score is then selected.
- Then, based on the observed selection results, update t (line 16) and n_i, t (line 22).

Algorithm 1: UCB Algorithm (Modified from original source [18] to the source code given at <https://github.com/KANG-XIAOHAN/Multi-Armed> accessed on 9 December 2021).

```

1  $t :=$  arm number;
2  $N :=$  total number of arms, number of rounds;
3  $T :=$  total time of playing arm, where  $T \geq N$ ;
4  $t_{round} :=$  the round of player test;
5  $reward :=$  the reward of each time play gambling;
6  $chosen :=$  the number of chosen arm;
7  $sc :=$  simplify for  $chosen$ ;
8  $ucbVal :=$  the UCB values estimate for each arm;
9  $\mu :=$  mean of distribution for each arm;
10  $PL \leftarrow play(t_{round})$ ;
11  $R \leftarrow random()$ ;
12  $INIT \leftarrow init()$ ;
13 if  $t$  to 10 then for  $chosen$  to 10 do
14    $chosen \leftarrow random()$ ;
15   if  $reward$  is max then  $chosen \leftarrow play(chosen, t_{round})$ ;
16    $t_{round} \leftarrow + 1$ ;
17  $INIT \leftarrow init()$ ;
18 for  $t$  to  $T$  do                                     // play arm  $t$  one by one,  $t \in N_+$ 
19    $chosen \leftarrow play(chosen, t_{round})$ ;
20   if  $chosen > 0$  then  $ucbVal = \hat{\mu}_{t,sc} + \sqrt{\frac{\ln sc}{n_{tsc}}}$ ;
21    $chosen \leftarrow play(reward + ucbVal)$ ;
22    $chosen \leftarrow update(t, chosen, reward)$  // depends on the  $reward$  and  $ucbVal$ ,
   we update the arm we choose for getting the maximum reward

```

3.3. Experiment Setup

The player energy changing over various masses were compared to clarify how a player feels engaged in the game process. Because of the data particularity, there is no such accurate open data for intuitive probability and actual probability for different mass values. In this study, 10 settings were simulated. The experiment took a random distribution conditional on m being selected from 0 to 1, where the details of the distribution for each m are given in Tables 3 and 4 for the 3-armed bandit and 10-armed bandit, respectively. Such an experiment was designed to separate the effects of each Multiarmed Bandit in a different mass by controlling each arm's distribution sets.

Table 3. Experiment setting for 3-armed bandit.

Arm Setting Distribution	m	Arm Numbers
(0,1)(0,1)(0,1)	0	3
(−1.03,1)(−1.22,1)(−1.75,1)	0.1	3
(−0.77,1)(−0.68,1)(−1.12,1)	0.2	3
(−0.14,1)(−0.51,1)(−0.99,1)	0.3	3
(−1.03,1)(−0.55,1)(0.71,1)	0.4	3
(0.30,1)(−0.56,1)(0.22,1)	0.5	3
(2.04,1)(0.20,1)(−0.71,1)	0.6	3
(0.61,1)(0.73,1)(0.25,1)	0.7	3
(4.13,1)(0.77,1)(0.31,1)	0.8	3
(4.16,1)(1.32,1)(0.80,1)	0.9	3
(4.27,1)(4.27,1)(4.27,1)	1.0	3

Table 4. Experiment setting for 10-armed bandit.

Arm Setting Distribution	<i>m</i>	Arm Numbers
(−4.5,1)(−4.5,1)(−4.5,1)(−4.5,1)(−4.5,1)(−4.5,1)(−4.5,1)(−4.5,1)(−4.5,1)(−4.5,1)	1.0	10
(−2.07,1)(−0.94,1)(−1.42,1)(−4.77,1)(−0.90,1)(−1.28,1)(−1.07,1)(−1.04,1)(−1.36,1)(−1.46,1)	0.9	10
(−0.43,1)(−0.78,1)(−3.46,1)(−1.01,1)(−0.75,1)(−0.65,1)(−1.21,1)(−1.22,1)(−0.47,1)(−0.59,1)	0.8	10
(−0.87,1)(0.20,1)(−0.80,1)(−0.86,1)(−1.07,1)(0.17,1)(−1.40,1)(−0.21,1)(0.19,1)(−1.62,1)	0.7	10
(−0.09,1)(0.75,1)(0.52,1)(1.36,1)(−0.83,1)(−1.53,1)(−2.22,1)(−0.58,1)(−1.18,1)(−0.09,1)	0.6	10
(−0.92,1)(1.13,1)(−0.80,1)(−0.82,1)(0.50,1)(0.19,1)(0.53,1)(0.78,1)(0.24,1)(−0.94,1)	0.5	10
(0.74,1)(0.82,1)(0.11,1)(0.17,1)(0.65,1)(0.06,1)(−0.55,1)(0.31,1)(−0.23,1)(0.62,1)	0.4	10
(−1.99,1)(−0.90,1)(−0.29,1)(−1.55,1)(−1.10,1)(−0.75,1)(−0.50,1)(−0.68,1)(−0.42,1)(−1.27,1)	0.3	10
(0.39,1)(0.69,1)(0.39,1)(1.77,1)(0.89,1)(1.60,1)(0.92,1)(0.79,1)(1.03,1)(0.73,1)	0.2	10
(2.18,1)(1.11,1)(1.91,1)(1.25,1)(1.76,1)(1.22,1)(0.53,1)(1.01,1)(1.33,1)(2.50,1)	0.1	10
(0,1)(0,1)(0,1)(0,1)(0,1)(0,1)(0,1)(0,1)(0,1)(0,1)	0	10

The Multiarmed Bandit in this simulation follows Gaussian distribution, where every arm follows the Gaussian distribution. For Bayesian, the probability of spending money at each slot machine has a prior distribution assumption as long as we enter the same casino. After pushing the slot machines, the corresponding posterior distribution can be adjusted according to the related feedback. There are 10 sets of experiments in this section that corresponding to different reward distributions. The simulated slot experiment aims to estimate the overall expectation of slot machines throwing money through the known sample distribution. It is a Bayesian process since each arm obeys the Gaussian distribution. Suppose that the component with a higher feedback rate among *n* arms can be found. In that case, the joint distribution of multiple Gaussian distributions needs to be analyzed, which is the binomial distribution process. Based on this, two sets of experiments with 3-armed and 10-armed were performed to analyze player psychology, and 11 groups of experiments were compared. Uncertainty of the game is controlled by setting up different reward distributions as shown in Table 3 and 4. There are 10,000 times training for each setting to simulate the selection process using the UCB method to maximize the next-choice reward. We collected data on predicted expectations before each choice and true rewards after each option, and then compared and analyzed them.

An example of the 3-armed bandit game was depicted in Figure 2, where it showed different game levels between the predicted reward and actual reward. The blue line shows the predicted reward, and the orange line shows the actual reward. The figure demonstrates the first 300 training results by using Savizky–Golay filter to less noisy.

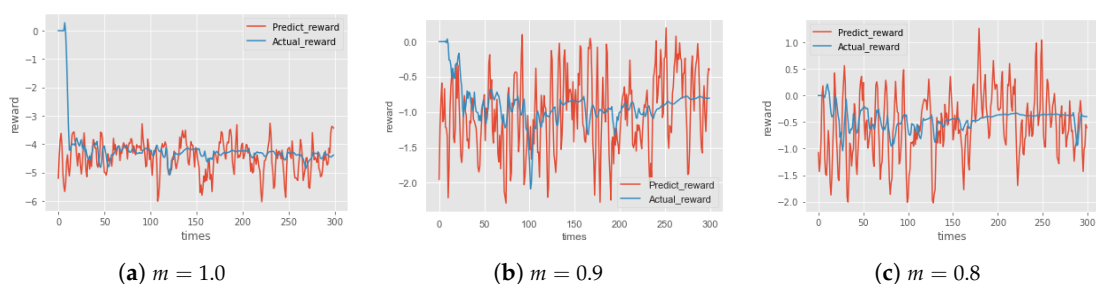


Figure 2. Cont.

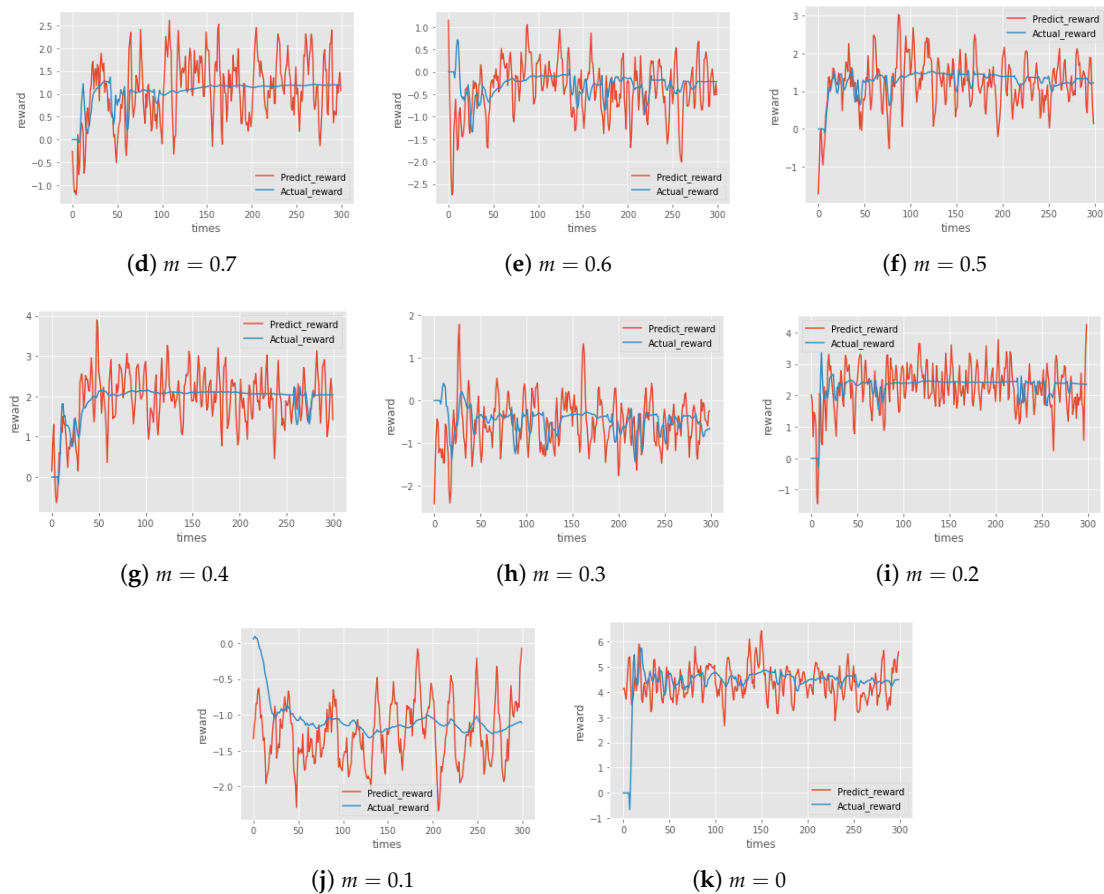


Figure 2. Comparison of predicted reward and actual reward with a game length of 300 steps with $m \in [0, 1]$ (m is mass in game).

4. Results and Analysis

In this paper, two sets of experiments with 3-armed and 10-armed bandits were performed to analyze player psychology, and 11 groups of experiments were compared. Uncertainty of the game is controlled by setting up different reward distributions as shown in Tables 3 and 4. There are 10,000 times training for each setting to simulate the selection process using the UCB method to maximize the next-choice reward. We collected data on predicted expectations before each choice and true rewards after each option, and we then compared and analyzed them.

4.1. Psychological Gap Expressed by Energy Difference

Higgins [31] proposed the theory of ego-fall, where he argues that the ideal-self and the real-self are the standards that guide the authentic-self to reach. When the gap between the real-self and the ideal-self is created, the motivation to reduce this gap arises, and this motivation drives behavior and makes people strive.

As is shown in Figure 2, the range of the predicted reward is larger than the actual reward. Furthermore, in the game length for each level, the range of the predicted reward is always more extended than the actual reward range, which indicates that player prediction is unstable. Therefore, there is always a difference between actuality and prediction. In other words, the player's perception of uncertainty fluctuates much more than the actual reward; thus, creating a psychological gap between prediction and reality while playing.

To differentiate the difference of the psychological gap between prediction and reality in gaming, energy difference E_d is computed and reported in Tables 5 and 6. There are two peaks as m increases, where $m = 0.3 - 0.4$ and $m = 0.6 - 0.7$. The energy difference can be up to 0.29504 and 0.24365 for two settings (Figure 3). When $m = 0.3 - 0.4$ and

$m = 0.6 - 0.7$, the player has the biggest psychological gap, which expresses that player will be motivated by not reconciled. The high psychological gap makes players think that they may win in the next pull which makes them continue to play. In this experiment, the energy difference is decreasing when m is decreasing since the uncertainty of the game is decreasing, which shows that the players gain more confident in their prediction. When $m = 0$ and $m = 1$, the energy difference is reaching to 0, which shows that the actual game results satisfied the player prediction.

It is an extreme case that no-lose or no-win would happen in the game, which is easy to predict. The compared energy difference between 10-armed bandit and 3-armed bandit shows that the energy difference is in a similar range. Moreover, there is always a sudden drop while $m = 0.5$, which shows that game-side energy is closer to player-side energy, the game is relatively fair. A 3-armed bandit expressed more unstable than 10-armed since it has fewer choices, while one judgment will reflect more than 10-armed bandit.

Table 5. Results of energy difference in 3-arm bandit.

m	Actual Probability	Intuitive Probability	Energy Difference E_d
1.0	0.00000	0.00000	0.00000
0.9	0.00001	0.08100	0.01206
0.8	0.00014	0.23491	0.08444
0.7	0.00150	0.34777	0.15776
0.6	0.33433	0.38396	0.03282
0.5	0.66537	0.50869	-0.04202
0.4	0.66776	0.59327	-0.00998
0.3	0.99994	0.69022	0.29504
0.2	0.99983	0.81218	0.24744
0.1	0.99995	0.89007	0.17408
0.0	0.99997	1.00000	-0.00001

Table 6. Results of energy difference in 10-arm bandit.

m	Actual Probability	Intuitive Probability	Energy Difference E_d
1.0	0.00000	0.00000	0.00000
0.9	0.00002	0.17683	0.05147
0.8	0.00013	0.30717	0.13074
0.7	0.99744	0.58729	0.27960
0.6	0.99973	0.88805	0.17603
0.5	0.99978	0.91979	0.13527
0.4	0.99962	0.81684	0.24365
0.3	0.00003	0.46733	0.23266
0.2	0.99989	0.96995	0.05632
0.1	0.99988	0.99971	-0.00033
0.0	0.99990	1.00000	-0.00019

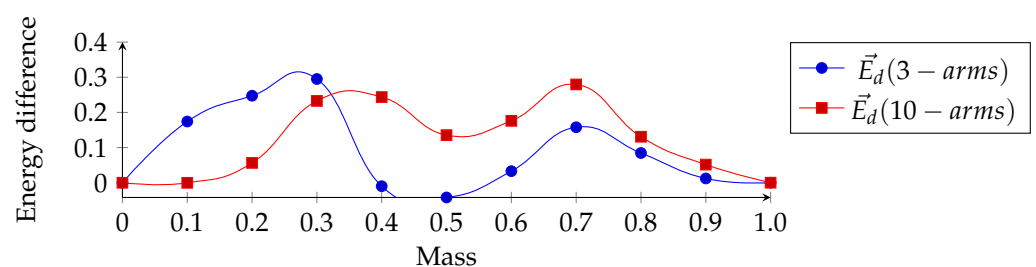


Figure 3. Changes of energy difference measures.

4.2. Link between Satisfaction and Competitive in Game Playing

Based on the previous study by Iida and Khalid [26], potential energy is ‘skewed’ towards a player with a sufficiently high (but not necessarily perfect) ability, while momentum is the greatest when the player possesses the ability similar to the majority of the players of such game. Momentum makes players more competitive to play [32], while energy determines whether or not players are satisfied with the game. In the moment where momentum equals energy ($\vec{p} = E_p$), player satisfaction and competitive feeling are well balanced (denoted as player motivated point). When $\vec{p} > E_p$, the player would be more competitive. Meanwhile, $\vec{p} < E_p$, the player would be more satisfied but less motivated.

Energy difference E_d provided the player psychology gap in-game process. As shown in Figure 4, in 3-armed bandit, when $\vec{p}_d = E_d$, $m \in [0.3, 0.4]$ and $m \in [0.6, 0.7]$. Meanwhile, for 10-armed bandit, $m \in [0.2, 0.3]$ and $m \in [0.7, 0.8]$ when $\vec{p}_d = E_d$. The range on both settings was closely similar, which can be associated with players who are well-motivated due to competition and satisfaction. Nevertheless, there were some limitations in light of this study’s findings. With the change of exact arm setting, the mass value will make subtle differences. The study results highlight the need for future research to use a representative sample.

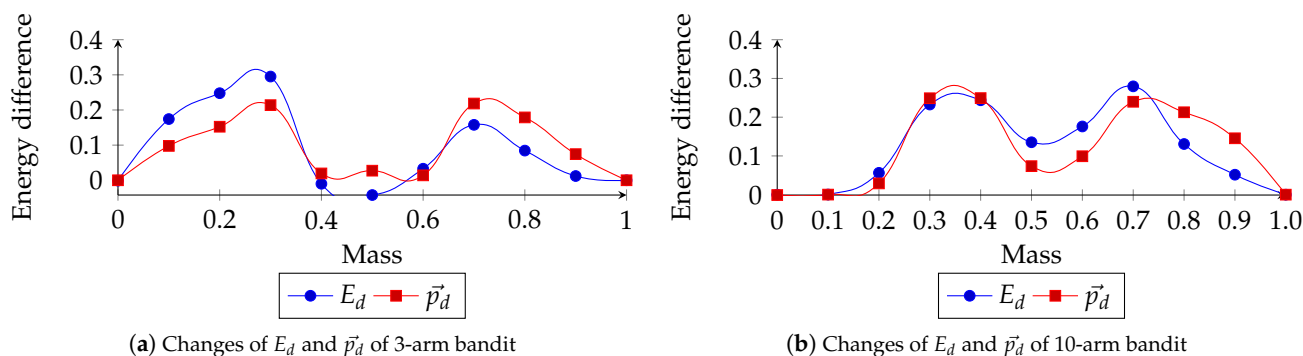


Figure 4. Changes of energy difference measures.

5. Discussion

5.1. Application with Player Fairness Domain

In the motion in mind model [26], $m = 0.5$ is the absolute middle-ground between fair and unfair. However, when $m > 0.5$, the play condition will favor the game side and become more competitive. In contrast, the play condition will favor the opposite (player side) when $m < 0.5$, which is associated with being more satisfied. As mentioned before, the player motivated point is around $m \in [0.3, 0.7]$, as shown in Figure 5. It can be conjectured that when $0.3 < m < 0.5$, the player would be more satisfied but less competitive; naturally, in the educational context, which needs more encouragement and less uncertainty. When $0.5 < m < 0.7$, the player would be more competitive but less satisfied, which appears in sports and competitive games.

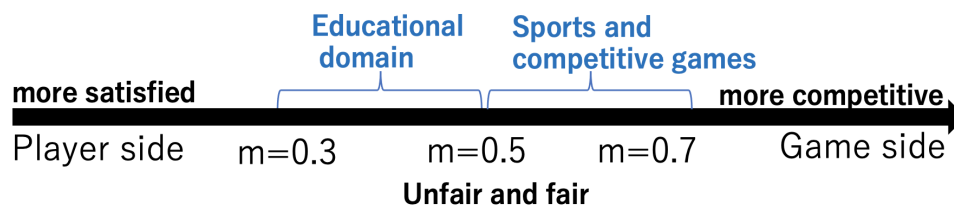


Figure 5. Application with player fairness domain.

5.2. Why Is the Multiarmed Bandit Addictive?

The physical excitement of gambling, the great joy, and the sadness caused by the substantial psychological gap between winning and losing bring pleasure to the body. Like

roller coasters and skydiving, it is difficult for other recreational activities to provide. The pursuit of this kind of exciting fun is the most direct, simple, and initial reason. The energy difference E_d provides the difference between prediction and actual reward to show the player psychology gap.

Secondly, the motivation which pushes a player to continue the game is to balance the psychology gap. When E_i is more extensive, the player side has more influence. On the contrary, the game side will influence more. To encourage the player to play the game, energy difference E_d will be positive for strengthening player confidence and reinforcing the reward effect. Furthermore, when \bar{p}_d equals E_d (player satisfaction and competitive feeling are well balanced), m lands up to around 0.3, 0.7 in the two settings of this study. Additionally, gambling needs to be considered to guarantee a profit while encouraging the game continuation, so m lands up to around 0.3, which can be evaluated in real gambling games.

Thirdly, energy difference E_d can be applied to many areas to analyze whether player confidence was motivated, such as educational areas and business models. It is suggested that the mass of such a game could be controlled in the range of $m \in [0.3, 0.5]$. The games that focus on competitive and thrilling feelings should be at the stage of $\in [0.5, 0.7]$. In essence, the mass value should always be $\in [0.3, 0.7]$ to fill the psychology balance.

Finally, the game is a process in which the player constantly tries to balance their psyche and make behavioral judgments through empirical evaluation. In this learning process, expectations and disparities shape the player's psychology. Expectations can be understood in the abstract related to challenges, and differences are formed mainly by the gap between reality and ability or between the opponent/game's side and the player's side. Therefore, a good game can help the player achieve a balance between psychological competition and satisfaction while encouraging and guiding the player to continue the game process and achieve psychological comfort. In the education sector, such gamification can be designed to facilitate learning planning and goal attainment.

5.3. Limitation

This paper selects one of gambling's multiarmed slot machines for study and analysis. The single nature of the game's reward mechanism makes it the best object of study to examine the psychology of players based on a reward system. The findings may be limited to application for the quantification of player psychology in the context of any randomized reward system. In addition, on an individual basis, this paper's methodology can also be limited to player segmentation. For example, players who maintain a solid willingness to continue playing when the energy difference is negative and consistently pessimistic can be called unbeatable players. Players who continue to play only when their energy is positive can be referred to as encouraging players.

6. Conclusions

This study identified the reward mechanism on Multiarmed Bandit games using the analogy of energy difference in games. Thus, the player's interactivity and games can express the psychological gap to understand motivation and possibly addiction better. This situation addresses our first objective to better clarify the psychological gap of players by mapping the reward of the Multiarmed Bandit games relative to the player satisfaction model [10]. Furthermore, it was found that the difference between intuitive and actual probability is where player motivation comes from, as denoted by the positive energy difference. Thus, high reward expectations, in spite of low actual returns, motivate players, while some negative energy difference causes the experience to be surprised and encouraged.

This study demonstrated that the game process could be a motivational tool for learning and entertainment, where players react differently regarding rewards and uncertainty. In addition, the measures of energy difference provide a quantification tool to better analyze the player psychology of the Multiarmed Bandit game by providing a controlled environment of uncertainty (based on the m value). Finally, based on the simulation results,

a balanced setting provided a fair and potentially motivating point (in contrast to addictive) that could be useful for learning and entertainment perspectives. These points highlighted the underlying mechanisms behind players' psychological inclination and possible reasons why gambling games are addictive; thus, achieving our second objective of the study.

Based on the energy difference (E_d) in Multiarmed Bandit games, it was found that a player's psychological gap can be computationally estimated to identify player confidence ($E_d > 0$) which encourages the player to continue gaming. In contrast, player frustration ($E_d < 0$) can also be identified, discouraging players due to entrance difficulty. Furthermore, considering the relations of the energy measures to the momentum (\vec{p}), the intersections between momentum difference and energy difference ($E_d = \vec{p}_d$) potentially describe the player's motivation point, which fulfills player satisfaction and the sense of competitiveness.

In essence, a game is a process where players constantly try to balance their psyche and judge their behavior through empirical evaluation, shaped by the expectation and disparities of their learning process. Thus, the challenge faced by the players is abstracted by their expectations. Meanwhile, the disparities were demonstrated based on the gap between the game element and the player psyche. As such, a well-designed game help players psychologically achieve a balance between competitiveness and satisfaction while encouraging and guiding the player to continue the gaming experience. Such a case would be beneficial in modeling educational and business processes concerning the concept of gamification [33], where learning in both contexts can be optimized while providing an enjoyable experience.

Author Contributions: Conceptualization, X.K., H.R., M.N.A.K. and I.H.; data curation, X.K. and H.R.; formal analysis, X.K., H.R., M.N.A.K. and H.I.; funding acquisition, H.I.; investigation, X.K. and H.R.; methodology, X.K., H.R., M.N.A.K. and H.I.; project administration, H.I.; resources, H.I.; software, X.K. and H.R.; supervision, H.I.; validation, X.K., H.R., M.N.A.K. and H.I.; visualization, X.K., H.R., M.N.A.K. and H.I.; writing—original draft, X.K., H.R., M.N.A.K. and H.I.; writing—review & editing, X.K., H.R., M.N.A.K. and H.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by a grant from the Support for Pioneering Research Initiated by the Next Generation Home, JST Spring (Grant Number: JPMJSP2102).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not Applicable, the study does not report any data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Skinner, B.F. *The Behavior of Organisms: And Experimental Analysis*; Appleton-Century-Crofts: New York, NY, USA, 1938; p. 457.
2. Pfau, J.G.; Everitt, B.J. The psychopharmacology of sexual behavior. *Psychopharmacology* **1995**, *65*, 743–58.
3. Demos, K.E.; Heatherton, T.F.; Kelley, W.M. Individual differences in nucleus accumbens activity to food and sexual images predict weight gain and sexual behavior. *J. Neurosci.* **2012**, *32*, 5549–5552. [[CrossRef](#)] [[PubMed](#)]
4. Balleine, B.W.; Daw, N.D.; O'Doherty, J.P. Multiple forms of value learning and the function of dopamine. In *Neuroeconomics*; Elsevier: Amsterdam, The Netherlands, 2009; pp. 367–387.
5. Flagel, S.B.; Clark, J.J.; Robinson, T.E.; Mayo, L.; Czuj, A.; Willuhn, I.; Akers, C.A.; Clinton, S.M.; Phillips, P.E.; Akil, H. A selective role for dopamine in stimulus—Reward learning. *Nature* **2011**, *469*, 53–57. [[CrossRef](#)] [[PubMed](#)]
6. Volkow, N.D.; Wang, G.J.; Fowler, J.S.; Tomasi, D.; Telang, F. Addiction: Beyond dopamine reward circuitry. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 15037–15042. [[CrossRef](#)] [[PubMed](#)]
7. Green, C.S.; Bavelier, D. Learning, attentional control, and action video games. *Curr. Biol.* **2012**, *22*, R197–R206. [[CrossRef](#)] [[PubMed](#)]
8. Bellis, M. A Brief History of Slot Machines, 2019. Available online: <https://www.thoughtco.com/history-of-slot-machines-liberty-bell-1992409> (accessed on 10 December 2021).
9. Iida, H.; Takeshita, N.; Yoshimura, J. A metric for entertainment of boardgames: Its implication for evolution of chess variants. In *Entertainment Computing*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 65–72.

10. Xiaohan, K.; Khalid, M.N.A.; Iida, H. Player Satisfaction Model and its Implication to Cultural Change. *IEEE Access* **2020**, *8*, 184375–184382. [CrossRef]
11. Malone, T.W. What makes things fun to learn? Heuristics for designing instructional computer games. In Proceedings of the 3rd ACM SIGSMALL Symposium and the First SIGPC Symposium on Small Systems, New York, NY, USA, 18–19 September 1980; pp. 162–169.
12. Bartle, R. Hearts, clubs, diamonds, spades: Players who suit MUDs. *J. Mud Res.* **1996**, *1*, 19.
13. Yee, N. Motivations of play in MMORPGs. 2005. Available online: <https://summit.sfu.ca/item/212> (accessed on 10 December 2021).
14. Multi-Armed Bandit Wiki. Available online: https://en.wikipedia.org/wiki/Multi-armed_bandit/ (accessed on 9 December 2021).
15. McCall, B.P. Multi-Armed Bandit Allocation Indices (JC Gittins). *SIAM Rev.* **1991**, *33*, 154. [CrossRef]
16. Berry, D.A.; Fristedt, B. *Bandit Problems: Sequential Allocation of Experiments (Monographs on Statistics and Applied Probability)*; Chapman and Hall: London, UK, 1985; Volume 5, p. 7.
17. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
18. Zhang, S. Reinforcement Learning: An Introduction. Available online: https://github.com/ShangtongZhang/reinforcement-learning-an-introduction/blob/master/chapter02/ten_armed_testbed.py (accessed on 9 December 2021).
19. Skinner, B.F. ‘Superstition’ in the pigeon. *J. Exp. Psychol.* **1948**, *38*, 168. [CrossRef] [PubMed]
20. Mcgonigal, J. Game Can Make a Better World. 2010. Available online: https://www.ted.com/talks/jane_mcgongal_gaming_can_make_a_better_world (accessed on 10 December 2021).
21. Wang, J. Research on mergers and acquisitions of the international hotel group—Take Marriott M & A Starwood as an example. In Proceedings of the 2018 3rd International Conference on Humanities Science, Management and Education Technology (HSMET 2018), Nanjing, China, 8–10 June 2018; Atlantis Press: Paris, France, pp. 743–746.
22. Kerr, J.; Slocum, J.W., Jr. Managing corporate culture through reward systems. *Acad. Manag. Perspect.* **2005**, *19*, 130–138. [CrossRef]
23. Iigaya, K.; Hauser, T.U.; Kurth-Nelson, Z.; O’Doherty, J.P.; Dayan, P.; Dolan, R.J. The value of what’s to come: Neural mechanisms coupling prediction error and the utility of anticipation. *Sci. Adv.* **2020**, *6*, eaba3828. [CrossRef] [PubMed]
24. Fiorillo, C.D.; Tobler, P.N.; Schultz, W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* **2003**, *299*, 1898–1902. [CrossRef] [PubMed]
25. Shizgal, P.; Arvanitogiannis, A. Gambling on dopamine. *Science* **2003**, *299*, 1856–1858. [CrossRef] [PubMed]
26. Iida, H.; Khalid, M.N.A. Using games to study law of motions in mind. *IEEE Access* **2020**, *8*, 138701–138709. [CrossRef]
27. Ferster, C.B.; Skinner, B.F. Schedules of Reinforcement. 1957. Available online: <https://psycnet.apa.org/doiLanding?doi=10.1037%2F10627-000> (accessed on 10 December 2021).
28. Carathéodory, C. Untersuchungen über die Grundlagen der Thermodynamik. *Math. Ann.* **1909**, *67*, 355–386. [CrossRef]
29. Lai, T.L.; Robbins, H. Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes. *Z. Wahrscheinlichkeitstheorie Verwandte Geb.* **1981**, *56*, 329–360. [CrossRef]
30. Lai, T.L.; Robbins, H. Asymptotically efficient adaptive allocation rules. *Adv. Appl. Math.* **1985**, *6*, 4–22. [CrossRef]
31. Higgins, E.T. Self-discrepancy: A theory relating self and affect. *Psychol. Rev.* **1987**, *94*, 319. [CrossRef] [PubMed]
32. Agarwal, S.; Khalid, M.N.A.; Iida, H. Game refinement theory: Paradigm shift from performance optimization to comfort in mind. *Entertain. Comput.* **2019**, *32*, 100314. [CrossRef]
33. Lewis, Z.H.; Swartz, M.C.; Lyons, E.J. What’s the point?: A review of reward systems implemented in gamification interventions. *Games Health J.* **2016**, *5*, 93–99. [CrossRef] [PubMed]