Preferred Difficulty Levels in Novel Computer-Based Tasks

Thesis presented to the Faculty of

Notre Dame de Namur University

In partial fulfillment of the requirements for the degree of

Masters of Science in Clinical Psychology

by

Steven Riley

Approved by:

Michele Haley, Ph.D., Chairperson

Matthew Harris, Ph.D., Chairperson

Helen Marlo, Ph.D., Second Reader

Abstract

This research attempted to quantify the preferred percentage correct (PPC) on three computer-based tasks (one verbal, one visual and one kinesthetic) for a sample of Internet users. PPC was defined as the point where the sample gave no net preference to changing the difficulty level after participants completed a series of ten trials with a task at a fixed difficulty. According to several theories, including Csikszentmihalyi’s theory of flow, the Yerkes-Dodson law, and Vygotsky’s zone of proximal development, this research investigated whether computer users preferred moderate levels of difficulty when encountering new tasks. Specifically, it investigated whether the PPC for these novel tasks would be between 80% and 95%. The sample included ten individuals of mean age 29.3 who were recruited through a post on Facebook. Nine of these participants spent at least five hours per week on average playing video games. Using regression analysis, the data yielded PPC values of 85.9%, 71.0%, and 39.7% on the verbal, visual and kinesthetic tasks, respectively. Two of these were outside of the hypothesized range. This report describes the pilot study; the full study will be published in Fall 2014.

**Table of Contents**

Abstract 2

Preferred Difficulty Levels in Novel Computer-Based Tasks 5

Fun 6

Three Models of Fun 6

Csikszentmihalyi’s Flow 6

Schmidhuber’s Formal Theory of Fun 8

Frustration, Arousal, Yerkes-Dodson, and Vygotsky’s Zone of Proximal Development 10

Possible Confounds 12

Demographics 12

Neuroticism 12

Learning Modalities 13

Research Aims and Hypotheses 13

Hypotheses Pertaining to Demographics 14

Hypotheses Pertaining to Neuroticism 15

Hypotheses Pertaining to Learning Modalities 15

Hypotheses Pertaining to Overall Data 16

Method 16

Participants and Design 16

Materials and Apparatus 18

Procedure 20

Analysis 23

Discussion 26

References 29

Figure captions?

Appendix captions?

Preferred Difficulty Levels in Novel Computer-Based Tasks

Video game players are motivated to give up large amounts of money and time in order to play games. Blockbuster video games can gross upwards of $1 billion (“Top 10 highest”, 2012). The average American gamer spends approximately $140 per year on games, and one in seven Americans spends at least 5 hours per week playing games (Takahasi, 2010). There is obvious financial incentive for video game companies to “corner the market” on fun. In pursuing the business of gamers, those companies that have been able to provide consistent entertainment to the video gaming public likely have learned something about how to craft their games to suit gamers’ desires for challenges.

Even some non-gaming businesses have begun to employ game-like paradigms in order to keep employees (Silverman, 2013) and customers motivated. Over the last few years, the technique of “gamification” has become mainstream (Amos, 2013) . “Gamification” (Deterding, Dixon, Khaled, & Nacke, 2011) is the technique of giving rewards usually found in games, such as points, badges, levels and achievements, contingent upon the performance of the user. For instance, Duolingo (www.duolingo.com) rewards a user with points for every completed exercise. If a user achieves a certain point total, that user’s level increases. When a user completes a block of lessons, a badge is awarded and a notification is posted to Facebook. The effects of gamification on motivation seem to be positive, but the results are highly dependent upon both the context and the user (Hamari, Koivisto, & Sarsa, 2014). This research sought to explicitly quantify some of the information that might be implicit in the designs of successful games and gamified systems. Specifically, it sought to answer the question “what is the level of difficulty preferred by users?”

**Fun**

What is meant by “fun”? This research took “fun” to be synonymous with enjoyment while executing a task. This paper will examine three different theories of fun and optimal experience and will demonstrate that they predict a similar phenomenon: the preferred difficulty level chosen by a user should be high enough to be interesting but not so high as to be impossible. This research investigated participants’ preferred levels of difficulty in terms of their percentage of correct responses on a task in order to determine how often individuals want to get items correct while doing a task. This information was used to inform conclusions as to how the behavior of an individual may change when given a task that is either too hard (participant makes too many mistakes) or that is too easy (participant makes too few mistakes).

The three theories of fun integrated into this study include one based on affect regulation, one based on learning, and one based on a combination of affect and learning. The model based on affect regulation will examine Csikszentmihalyi’s “flow” theory (Csikszentmihalyi, 1990) and what it predicts about behavior. The model based on learning will look at Schmidhuber’s formal theory of creativity, fun, and intrinsic motivation (Schmidhuber, 2010). The model that combines the two is based on Vygotsky’s theory of zone of proximal development (Vygotsky, 1978) and the Yerkes-Dodson law (Yerkes & Dodson, 1908).

**Three Models of Fun**

**Csikszentmihalyi’s Flow.** Csikszentmihalyi’s conception of enjoyment for an individual engaging in a task includes the following eight factors: one, the task can be completed; two, the individual can concentrate on the task; three, the task has clear goals; four, the task gives immediate feedback; five, while doing the task, the individual’s attention focuses to the point of forgetting the stresses of daily life; six, the individual feels a sense of control during the task; seven, ego and self-concern disappear; and eight, the sense of duration of time alters (Csikszentmihalyi, 1990, p.49). Not all eight factors must be present for enjoyment to occur. When the majority of these factors are present, one can be said to be in a state of flow.

One of the corollaries of these conditions is that entering a flow state requires there to be a balance between the perceived difficulty of the task and the participant’s perceived skills (Csikszentmihalyi, Abuhamdeh, Nakamura, Flow & Dweck, 2005). Thus, flow states require the difficulty of a task to closely match the participant’s preferred difficulty level for that task. If the task is too easy, boredom results. If it is too hard, anxiety results. In between areas of anxiety and boredom is a thin channel that represents the possibility for flow. Figure 1 is a visual depiction of this flow channel (Csikszentmihalyi, 1990, p. 74).

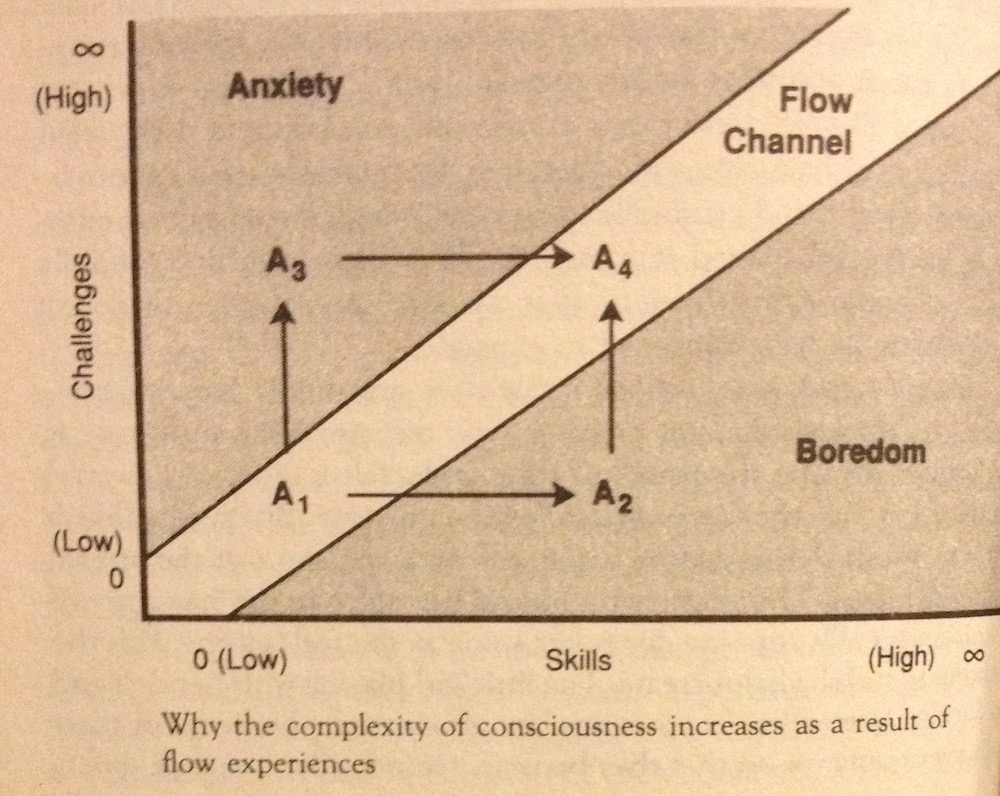


Figure 1: Flow Channel

As Csikszentmihalyi describes, the progression of skills of a participant repeatedly completing a task will move from A1 to A4. An increase in skills in the absence of an increase in difficulty leads to boredom (A2), which compels the participant to seek harder challenges. Increased challenges in the absence of increased skills causes the participant to practice more (A3), leading to increased skills. However, both of these paths assume that the participant is unable to choose the difficulty level of the challenge. According to the flow theory, if a participant could choose, the choice would be for a position directly in the middle of the flow channel.

A study showed that tasks that are too easy are more enjoyable than tasks that are too hard (Haworth, & Stephen, 1995), but that participants desired a good match between skills and challenges. Thus, this research predicted that participants would enjoy tasks with a difficulty level where the majority of tasks would yield successful results. However, one of the conclusions of the flow theory is that task participants want challenges. If a participant wanted to get 100% correct on a task, an optimal strategy would be to decrease the difficulty to its lowest point. But, the theory of flow predicts that this will not happen, and thus players will risk getting some percent of the tasks wrong in order to maintain a sense of enjoyment. This constrained the hypothesized range of preferred percent correct to high values less than 100%. This research estimated the range of preferred percent correct to be 80% to 95% for optimal fun.

**Schmidhuber’s formal theory of fun.** Schmidhuber (2010) proposed a model by which an understanding of the underpinnings of motivation may contribute to a theory of learning. . If an agent (human or artificial) has an internal model of the world, it requires a certain amount of memory to hold all of the information pertaining to that model. Once in a while a discovery might take place that would illuminate a redundancy in the data. This illuminating discovery would allow the agent to recode the information for the world model into a smaller footprint. For instance, Newton’s discovery of the law of gravity gave a common explanation for apples falling from trees and planets orbiting the sun (Schmidhuber, 2009).

Schmidhuber defines beauty as the compressibility of a data set (Schmidhuber, 2007). He claims that the beauty we see in a face, a poem, a sunset or a theory all stem from the same property - something about the observation allows us to store a large amount of information with a small amount of memory. Hudson (2011) tested this connection by comparing compression ratios of pieces of music that had been run through various lossless compression algorithms. He found that those pieces that were of a style considered typically “beautiful” (such as choral and orchestral masterpieces) where more highly compressible than pop, rock, techno and random noise.

Schmidhuber goes on to define interestingness, novelty, surprise, or fun as the change with respect to time of the required memory footprint for storing information about an object or idea. In other words, something is fun or interesting if it allows an agent to see beauty in a new place (i.e. to compress data which previously seemed incompressible). Well before Schmidhuber developed his theory, Davis (1971) ran a study attempting to figure out which theories people found “interesting”. Davis concluded that uninteresting theories confirmed an assumption of the participant, but interesting theories denied an assumption. Meaning, those ideas which participants found interesting were ones that allowed participants to recode their views of the world.

Schmidhuber says that in order to find something interesting (i.e. compressible), the observations an agent makes about the world must be neither predictably regular nor random. Instead, data from such observations must be “...regular in a way that is new with respect to the observer's current knowledge, yet learnable.” What does this say about an agent’s preferred difficulty level for a novel task? If fun is proportional to the ability to compress observational data, then in order to increase fun it is necessary to increase the rate at which an agent learns how to incorporate data into a new compression scheme. Meaning, the measure of fun is the rate at which the agent learns. If a series of tasks is so easy as to not require an agent to extend its capabilities, then that set of tasks will not be fun for the agent. Similarly, if the agent has so little knowledge or ability that the outcomes of a task seem random, there will be no compression of data and therefore no fun. Schmidhuber’s theory of fun leads to the conclusion that a moderate level of difficulty is ideal to produce the maximal amount of fun.

**Frustration, arousal, Yerkes-Dodson, and Vygotsky’s zone of proximal development.** While the idea of frustration might normally have negative associations, this paper will now argue that frustration can be an important aspect of fun. What is frustration? Dollard, Miller, Doob, Mowrer and Sears (1939) define frustration as the result of “an interference with the occurrence of an instigated goal-response”. In other words, frustration is the result of failure at a task at which one wanted to succeed.

Studies show that physiological arousal increases with frustration. For instance, Ivory and Kalyanaraman (2007) found that while playing video games, participants’ frustration correlated well with physiological arousal levels even when the games were non-violent.

The Yerkes-Dodson law states that as arousal increases, the ability of an organism to complete a difficult task initially increases, then hits a maximum, then decreases (Yerkes & Dodson, 1908). This relationship is commonly represented with an inverted *U*-shaped curve with “arousal” on the *x*-axis and “ability” or “performance” on the *y*-axis, as shown in Figure 2. The Yerkes-Dodson law has been applied in numerous environments, including predicting which obese children will lose weight (Johnston, Moreno, Regas, Tyler & Foreyt, 2012), describing how cognitive abilities vary in response to psychostimulants (Wood, Sage, Shuman & Anagnostaras, 2014), and explaining how the validity of witness testimony changes with the nature of the crime (Roberts, 2014).

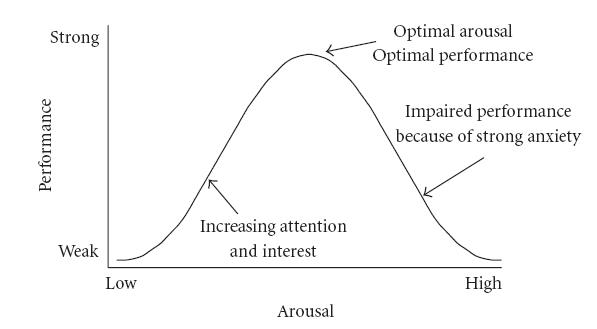


Figure 2, Yerkes-Dodson Law

As argued above in the section on Schmidhuber’s formal theory of fun, the interestingness of a task is related to the speed at which a person can develop a method of compressing information about the task (i.e., the speed at which the agent learns the task). According to Vygotsky’s idea of the zone of proximal development, learning takes place most effectively at the border between where an individual can complete a task on his or her own and where the individual cannot complete the task (Vygotsky, 1978). Since an individual will have his or her greatest level of performance on a task at a middling level of arousal, it follows that the difficulty level where the highest rate of learning would occur is where the amount of arousal generated equals this middling level. Thus, finding the difficulty level that corresponds to the peak on the Yerkes-Dodson curve amounts to finding the highest amount of fun for a task.

Since frustration can increase arousal levels, it follows that having some small amount of frustration can lead to higher performance, and therefore more fun. Meaning, a certain amount of frustration stemming from getting some percentage of items wrong on a task will lead to more fun than getting all the items right or all the items wrong. Too many items wrong increases frustration (and hence arousal) past the level of maximum ability on a task; too few wrong leads to decreased arousal below the peak. In conclusion, the relationship between frustration, arousal, the Yerkes-Dodson law and Vygotsky’s zone of proximal development leads to the prediction that there exists a non-zero percentage of wrong answers that will lead to the most fun.

**Possible Confounds**

**Demographics.** There are three main reasons why this research expected the demographics of the sample to be biased. First, participants accessed the study via a webpage, which would likely cause the sample to skew older and more affluent (Zickhur & Smith, 2013), and to include fewer minorities than would be expected (Prieger & Hu, 2008). Second, the primary method of participant recruitment was through online social networks, specifically Reddit, which would cause the sample to skew younger and more male (Duggan & Smith, 2014). Third, the study involved tasks that were similar to computer games, which would likely bias the sample towards males (Ogletree & Drake, 2007). In order to account for these possible selection biases, participants were asked for their age, gender, household income, race/ethnicity, and video game usage.

**Neuroticism.** This research cited frustration response and avoidance of negative stimuli in its theoretical justification. Thus, it was important to check for individual response to these two factors to see how much of an effect they would have on participant choices. According to Jeroniumus, Riese, Sanderman & Ormel (2014), “neuroticism is defined as the propensity to experience distress and negative emotions”, making neuroticism a good proxy for reactivity to negative stimuli. Additionally, frustration tolerance is inversely correlated with the neuroticism trait (Kundu & Basu, 1991). Thus, assessing for the neuroticism trait in participants was an appropriate way to control for both of these factors.

**Learning Modalities.** Since this research explicitly used learning as part of its theoretical justification, it was important to make sure that any learning related confounds were minimized. Specifically, slower learning might have increased the rate of getting answers wrong on a task. This would likely have increased frustration levels and decreased the level of difficulty with which a user felt most comfortable.

A possible factor affecting learning was the participant’s preferred learning modality, which is the mode through which the learner most readily absorbs information. This paper adopted the set of learning modalities in the Swassing Barbe Modality Index (SBMI) (Barbe & Swassing, 1979). These learning modalities are “visual”, “auditory” and “kinesthetic”. There is evidence that learning rate is independent of specific types of performance in certain settings. For instance, a study found that memory performance in a sensory domain is not correlated with learning modality (Krätzig & Arbuthnott, 2006). It also appears that several of the more popular learning style inventories disagree and are in fact measuring different things. In a 1997 classroom study, there was very little correlation between outcomes on four commonly used learning style inventories (Westman, Alliston, & Theriault, 1997). However, out of an abundance of caution, this research provided one game from each of the three SBMI learning domains in order to examine the possible effects of learning style on learning rate.

**Research Aims and Hypotheses**

The main purpose of this study was to quantify the percentage correct that corresponded to peak enjoyment. This preferred percentage correct could not have been found directly, so it was inferred from participant choices. There were two primary measures of preferred percent correct. The first was a measure of individual preferred percent correct (IPPC), and the second was an aggregate preferred percent correct (APPC) that could be determined for any subset of the sample. These two measures were computed using different methods but were intended to measure the same value. This value in all cases was expected to be in the 80% to 95% range.

**Hypotheses pertaining to demographics.** Due to the methods of data collection and participant recruitment, some skew in the sample was expected. Hypotheses one through five describe the expected distribution of demographic results in the sample. Hypotheses six through ten describe the expected outcomes on the preferred percent correct scores of the high and low game usage groups.

***Hypothesis 1.*** The sample will contain significantly more males than females.

***Hypothesis 2.*** The sample will be younger on average than the population as a whole.

***Hypothesis 3.*** The sample will include a higher percentage of Whites than the population as a whole.

***Hypothesis 4.*** The sample will be more affluent than the population as a whole.

***Hypothesis 5***. The sample will contain a higher percentage of heavy video game users than the population as a whole.

***Hypothesis 6.*** There will be no relationship between game-usage and IPPC.

***Hypothesis 7.*** There will be no difference in APPC between the high-game-usage and low-game-usage groups.

***Hypothesis 8.*** For each game usage group, there will be no difference in IPPC and APPC.

***Hypothesis 9.*** Mean IPPC score for each game-usage group will be in the 80%-95% range.

***Hypothesis 10.*** APPC score for each game-usage group will be in the 80%-95% range.

**Hypotheses pertaining to neuroticism.** Neuroticism was expected to relate to frustration tolerance and aversion to negative stimuli. Because of increased desire to escape frustration and negative stimuli, this study expected there to be a significant increase in preferred percent correct for those individuals with higher neuroticism scores.

***Hypothesis 11.*** Higher levels of neuroticism will correlate with higher IPPC.

***Hypothesis 12.*** The set of participants with the highest neuroticism scores will produce a higher APPC than the set with the lowest neuroticism scores.

**Hypotheses pertaining to task modalities.** Three task modalities were included in the study. Due to the relatively weak link in literature between preferred learning modality and performance on modality-related tasks, this study expected no relationship between task type and preferred percent correct.

***Hypothesis 13.*** There will be no significant difference between IPPC scores for the verbal, visual, and kinesthetic tasks.

***Hypothesis 14.*** There will be no significant difference between APPC scores of the data sets for verbal, visual, and kinesthetic tasks.

***Hypothesis 15.*** For each task modality, there will be no significant difference between IPPC and APPC scores.

***Hypothesis 16.***  Mean IPPC score for each task will be in the 80%-95% range.

***Hypothesis 17.*** APPC score for each task will be in the 80%-95% range.

**Hypotheses pertaining to overall data.** The APPC and IPPC scores are computed in very different ways, but should have been measuring the same value. This study expected that these numbers would be similar over the entire data set.

***Hypothesis 18.*** There will be no significant difference between the IPPC and APPC scores for the total sample set.

***Hypothesis 19.*** Overall mean IPPC score will be in the 80%-95% range.

***Hypothesis 20.*** APPC score for the entire data set will be in the 80%-95% range.

**Method**

**Participants and design**

Participants were advertised for online at Facebook. The flyer emphasized that the study involved playing three video-game-like tasks that would take approximately 15 minutes total, and that individuals with all levels of gaming experience are needed for the study (please see Appendix F). Would-be participants were directed to the website [www.ndnuvideogamestudy.com](http://www.ndnuvideogamestudy.com) in order to be part of the experiment.

Between 3/18/2014 and 3/25/2014, 11 participants took part in the study. Of these, data from the first 10 were included in the pilot data. The group was young (age range 18 - 36, *μ* = 29.3), and overwhelmingly consisted of people who spent large amounts of time gaming (*n* = 9 in the “high” game usage segment, *n* = 1 in the “medium” game usage segment, as described below). This split is vastly different from the general public, where approximately 14% of people qualify for the “high” game usage segment (“New Report”, 2013). The average age of video game users is 31 (Entertainment Software Association, 2014), which is close to the average age of the sample.

Participants were divided into three groups based on the amount of experience they had with computer gaming. These levels were called “low”, “medium” and “high” game usage. In order to define these groups, we relied on the definitions of The NPD Group (https://www.npd.com/wps/portal/npd/us/home/), a consumer market research firm that studies the gaming market. NPD research classified a “core” gamer as anyone who played video games for five or more hours per week (“New Report”, 2013). According to The NPD Group’s market research, this constitutes about 14% of the population (“New Report”, 2013). The current study adopted this definition of a “core” gamer to represent the “high” game usage segment. We defined the “low” usage group to include those individuals who were unlikely to play games in any given week. Correspondingly, the “medium” group consisted of those who played somewhere between zero and five hours of games during a typical week. Volunteers were sorted into “high”, “medium” and “low” game usage groups based on this definition. In order to sort participants into these categories, we asked participants to answer two Likert-scale questions about game usage habits as described in the “procedure” section below.

The design of this study was a 3x3 mixed-subjects quasi-experiment. The quasi-independent variables were “game usage” and “task type”. Levels for game usage were “low”, “medium” and “high”, as described above. Task type levels were “visual”, “verbal” and “kinesthetic”, as described in the materials and apparatus section below. Every participant fell into exactly one of the three game usage levels, so game usage was a between-subjects variable. Every participant completed all task types, so task type was a within-subjects variable. For each task type, there were two dependent variables. The first was “score at chosen difficulty”, which is a percentage from 0 to 100%. The second was “chosen difficulty after the test”. These two variables are described below.

For a given participant and a given task, the participant was asked to complete least 20 trials of that task across three different phases. Phase one was the tutorial, where the participant completed five tasks of increasing difficulty with the help of detailed instructions. Phase two was practice, where the participant was able to set the difficulty level before each task. The participant was able to end this phase whenever he or she wished. Before phase three started, the participant was asked to set the difficulty to the level he or she discovered to be the most enjoyable during the second phase. The difficulty was fixed at this chosen level during phase three. Phase three was the test phase, where the participant completed 15 trials at this self-selected difficulty level. Each trial from the second and third phases generated a data point (*x* = difficulty level of task, *y* = correct/incorrect response).

When the participant completed the third phase, he or she was asked: “if you were to do this task again, what difficulty would you prefer?” The difference between this value and the previously expressed preferred difficulty value was recorded as the second dependent variable described above.

**Materials and apparatus**

Participants accessed the three tasks on a computer. These tasks were the visual, verbal, and kinesthetic tasks described below. Each task was introduced by a short tutorial describing how to input answers and adjust the difficulty level. Scoring and score reporting were automated.

Each of these tasks had to fulfill the following characteristics in order to work for the design of the current experiment. One, it must be clear when there had been a success or failure on the task. Two, there must have been some easily adjustable component of the task which scaled either directly or inversely with difficulty. Three, the task must have been short enough so that 20 or more repetitions could have been completed in a reasonably short period of time (five minutes or less, ideally). Four, the task must have been novel but quickly learnable, meaning that it should have been composed of activities which were familiar to the participant but which were combined in manner with which the participant was not familiar. Five, answers must have been able to be inputted with a short series of keystrokes or mouse movements.

For the visual task, the participant was presented with two triangles. The participant was asked to visualize the circles that circumscribed each triangle. The participant then decided whether these two circles were the same size by means of a series of buttons marked “left is bigger”, “same size”, and “right is bigger” (Please see Appendix A for example). The participant had no time limit for this task. During all phases, after the participant selected an answer, the display revealed the circles and their relative sizes (Please see Appendix B for example). One third of the time the circumscribed circles were the same size. Two thirds of the time the circles were of different sizes, which were equally split between having the larger circle on the left and having it on the right. On items where the circles were not the same size, the closer these two circles were to the same size, the harder it was to discriminate between equal-sized and unequal-sized pairs. When the difference in size between the circumscribed triangles fell below the participant’s discrimination threshold, the participant should have been no better than random chance at deciding between similar and different sizes. The slider associated with this task adjusted the difference in radius of pairs of unequal size. The range of values on the slider was “0” to “50”, measured in pixels of radius difference.

For the verbal task, the participant was presented with a target word and then a list. The participant had four seconds to determine how many of the words in the list rhymed with the target word. The list contained words which were randomly chosen from one of four categories: rhyming lookalikes (“our” and “sour”); non-rhyming lookalikes (“our” and “pour”); rhyming non-lookalikes (“our” and “flower”); and non-rhyming non-lookalikes (“our” and “grower”) (please see Appendix C for example). When the participant was prompted with the word list, the clock started ticking down from four. The participant was instructed to click on a button indicating the correct number of rhyming words in the list. If the timer hit zero before the participant typed an answer, the target word and the word list disappeared, and were replaced by a prompt to choose an answer. When the participant entered an answer, the correct set of answers was revealed (please see Appendix D for example). The more words in the list, the harder it was to complete the task in four seconds. The slider in this task adjusted the number of words in the list. The slider’s range extended from one word to ten words.

For the kinesthetic task, participants were asked to use the arrow keys to guide a ball through a randomly generated maze to a blue end square (Please see Appendix E for example). Success meant getting the ball to the square without hitting a wall; if the participant collided with a wall, it was considered a failure. When a wall was hit, the ball disappeared and the wall blinked red where the ball hit the wall, and then the ball reappeared at the beginning of the maze. After a success, a new maze was generated. The slider adjusted the radius of the ball from 1 pixel to the half the width of a corridor, roughly 60 pixels.

**Procedure**

Before this study began, a proposal was submitted to the Notre Dame de Namur University Institutional Review Board (IRB). Drafts of this proposal were given to the second reader, Dr. Helen Marlo, and the thesis advisor, Dr. Nusha Askari. The thesis advisor, the second reader, and the IRB each signed off on the study before any data were collected.

Ads were placed on Facebook for volunteers. The ad emphasized that individuals who were 18 years or older with all levels of gaming experience were needed, and that the study involved playing games on a computer for approximately 10 to 15 minutes. Please see Appendix F for a copy of the ad.

In order to be enrolled in the study, prospective participants visited a website (www.ndnuvideogamestudy.com) containing a program which automated the study and sent data back to a server. Both the program’s code and the server code will be made publicly available at the end of the study. The program removed identifying information from all data sent over the Internet. The program and its associated database were hosted on a public web server. No information that could be used to identify participants was or will be stored on the server.

The participant first encountered an age-verification prompt, and then an informed consent form. In order to proceed, the participant had to give a birthday consistent with an age of 18 or older, and then had to acknowledge informed consent. Failure to do either caused the program to terminate. Please see Appendix G for a copy of the informed consent form.

Once age was verified and informed consent was given, the participant answered a two-question survey about gaming habits. The first question was “in a given week, how likely are you to spend any time playing video games?” The participant was presented with a Likert scale where one = “very unlikely”, two = “unlikely”, three = “equally likely as not”, four = “likely”, and five = “very likely”. If the participant answered with a one or a two, that participant was sorted into the low game usage category as described above. If the participant responded with a three, four or five, that participant was asked a follow-up question: “in a given week, how likely are you to spend AT LEAST FIVE HOURS playing video games?” Another Likert scale with the same ratings appeared, and the participant indicated a response. An answer of one or two put the participant in the medium game usage category, and a response of three through five put the participant in the high game usage category.

Once gaming habits were recorded, the program guided the participant through the three tasks described above in the materials section. Task order was completely counterbalanced and randomized. That is, there was an equal chance of the participant being asked to complete the tasks in any of the six possible orders (*ABC*, *ACB*, *BAC*, *BCA*, *CAB*, and *CBA*, where *A* = verbal, *B* = visual, and *C* = kinesthetic).

At the beginning of each task, there was a short (less than one minute) tutorial on the nature of the task, how to input answers, and how to alter the difficulty level via the slider. Participants were guided through at least 20 items of each task. The first five items comprised the tutorial. Between the tutorial and the test, the user was able to set the difficulty level on each item and practice for as long as he or she wished. After the participant chose to end the practice session, he or she was prompted to pick the level of difficulty level that was the most fun. Following this selection was the test session, which was composed of 15 items at this chosen difficulty level. At the end of each task, the participant was asked, “if you were to do this task again, what difficulty level would you want it to be?” The answer to this question was recorded and used for the analysis described below.

Once all three tasks were completed, the responses and any additional comments were coded into an HTTP-post request to hidden page on the [www.ndnuvideogamestudy.com](http://www.ndnuvideogamestudy.com) website. The page that caught the request (and stored the data in a MySQL database) was written in PHP. The program itself was written in Java, digitally signed using a code signing certificate from Comodo, and deployed from the website using Java Web Start. Code for the program, the Java Web Start launcher, the HTTP-post catcher, and the database extraction (also written in PHP) will be available online after the completion of the data gathering phase and closure of the database.

**Analysis**

The main analysis of this experiment focused on establishing estimated preferred difficulty values for each of the nine combinations of task type and game usage. For each participant, three estimates of this measure were possible from the data. Two are described here, and one is described in the secondary analyses section. The first estimate of preferred difficulty is the score that the participant received on the fifteen test items of each task. The reason that this is an estimate of preferred difficulty is that the user chose the difficulty level as the most fun. The second estimate can be interpolated from the combination of percentage correct on the task and the desired change in difficulty level after the task is complete.

The first estimate is a measure of prospective preferred difficulty. That is, it is a measure of what a participant would like for the near future. This can be calculated by a simple percentage of correct answers for the test phase of the task. Each participant generated one such percentage for each task. Averages per user varied from a minimum of .467 to a maximum of .933, with a mean of .723 and a standard deviation of .129. The Kinesthetic task produced the lowest average at .61, and the Verbal task produced the highest average at .8. The mean average of a task was .723 with a standard deviation of .100. These data deviate significantly from the hypothesized value of 80% to 95%. Please see Figure 14 for this data.

The second estimate was a measure of reflective preferred difficulty. That is, it measured the participant’s attitude toward his or her experience with the recent past. For each task, a participant generated a score and a preferred change in difficulty, or Δdiff, which was described in the method section. These pairs generated a scatter plot with score on the *x*-axis and Δdiff on the *y*-axis. These Δdiff scores were only comparable for a given task type. By looking at different subsets of the data, we can generate twelve such scatter plots. These are: nine for each cell in the 3x3 grid of data; and three for the total set of data in each of the three task types. However, given the skew in game usage type found in the pilot study, this analysis will only focus on the total data set for each task type.

For each scatter plot, a regression can be performed. According to the theories set out in the introduction, higher scores on tasks should lead to desires for increased difficulty (positive Δdiff), and lower scores will lead to desires for decreased difficulty (negative Δdiff). Thus, we would expect there to be a negative slope to the regression line. (Since for the Kinesthetic and Visual tasks, high slider values indicated easier tasks, we would expect this result to be reversed on those tasks). We would expect the *x*-intercept of this line to fall somewhere between the minimum and maximum score for the task. This *x*-intercept represents the best estimate of the population’s desired difficulty level. That is, this point on the regression line is where the predicted Δdiff equals zero, and where a participant would likely want to keep the difficulty the same.

In order to test the hypothesis that lower scores led to higher Δdiff values on the Verbal task, a regression was performed. The result approached significance, with *β* = .604, *R*2 = .364, *p* = .065, *p* > .05. In order to test the hypothesis that lower scores led to lower Δdiff values on the Kinesthetic task, a regression was performed. The result was not significant, with *β* = -.322, *R*2 = .104, *p* = .364, *p* > .05. In order to test the hypothesis that lower scores led to lower Δdiff values on the Visual task, a regression was performed. The result approached significance, with *β* = -.584, *R*2 = .341, *p* = .076, *p* > .05.

To test whether the *x*-intercept of the regression line on the Verbal task was within the hypothesized range, the regression line was constructed. The equation of this regression line was *y* = -2.97 + 3.46*x*. The *x*-intercept had a coordinate of .859, which was within the hypothesized range of .8 to .95. To test whether the *x*-intercept of the regression line on the kinesthetic task was within the hypothesized range, the regression line was constructed. The equation of this regression line was *y* = 2.2 - 5.54*x*. The *x*-intercept had a coordinate of .397, which was outside of the hypothesized range of .8 to .95. To test whether the *x*-intercept of the regression line on the Visual task was within the hypothesized range, the regression line was constructed. The equation of this regression line was *y* = 5.6 - 7.89*x*. The *x*-intercept had a coordinate of .710, which was not within the hypothesized range of .8 to .95.

To test the hypothesis that there was no difference between preferred difficulty levels for the three different tasks, within-subjects, repeated measures ANOVA test was run on the outcome percentages of each participant. For this ANOVA, *F*(2, 8) = 2.174, *p* = .176, *p* > .05. This result did not achieve significance, which was in agreement with hypothesis four.

Another question of importance for this study is whether learning continued to take place during the test session, and if so, at what rate? Ideally no learning would take place, since that could confound the relationship between percentage correct and Δdiff as described above. In order to test for the presence of learning, a linear regression was run on a scatter plot of percentage correct vs. item number on test phase data. If there was learning taking place, then one would expect the slope of the linear regression to be significantly greater than zero.

To test whether learning occurred in the kinesthetic task, a plot of index vs. outcome was generated for all kinesthetic test data, and a regression was run. No significant relationship was found, with *β*= .003, *R*2 = .000, *p* = .969, *p* > .05. To test whether learning occurred in the Visual task, a plot of index vs. outcome was generated for all Visual test data, and a regression was run. No significant relationship was found, with *β*= -.051, *R*2 = .003, *p* = .539, *p* > .05. To test whether learning occurred in the Verbal task, a plot of index vs. outcome was generated for all Verbal test data, and a regression was run. No significant relationship was found, with *β* = -.322, *R*2 = .104, *p* = .371, *p* > .05.

**Discussion**

Though data was sparse for this experiment, there were some interesting patterns that emerged. First of all, since the *p*-values for the beta coefficients approached significance in two of the three task types, it looks as though there is a relationship between percent correct and a user’s change in difficulty, lending some support to hypotheses one and two. This result agrees with all of the theories laid out in the introduction section. In the full version of this experiment, we are hopeful that there will be significant values for these beta coefficients, which would indicate strong support of the hypothesis and provide experimental evidence for the flow theory, the formal theory of fun, and the zone of proximal development theory.

Second, there seems to be a lack of evidence for learning taking place after the beginning of the test section. Once more data is available, we will run more types of tests to be sure. These tests will include logistic as well as linear regressions, and generation of *a posteriori* logistic curves using bootstrapping.

Third, it appears that we overestimated the percentage correct that participants preferred on tasks. Few participants scored above 80% at their preferred difficulty level, and the average percent correct of 72.3% is distinctly lower than the 80% - 95% hypothesized range. Additionally, the regressions of percent correct vs. Δdiff intersected the *x*-axes at lower values than expected. The *x*-value of 39.7% for the kinesthetic task was especially noteworthy.

It was intended that the data from the practice phase of each trial would be used to attempt to construct a logistic function relating difficulty level to percentage correct for each participant using the maximum *a posteriori* method. This method uses Bayesian reasoning to establish parameters of a distribution that generate the highest likelihood of a given set of observations. For each item in phase one, a participant will have a difficulty level and either a success or failure (coded as a 1 or a 0). As described in the first part of the introduction, it is assumed that the function relating difficulty level and chance of success is logistic. Any logistic function can be completely described by a mean and a shape parameter. For each combination of mean and shape parameter, it is possible to generate a likelihood that the participant’s responses would be the given set of successes and failures. By examining this likelihood function over the set of all means and shape parameters, it might have been possible through search methods to find it and use it to predict how hard a task would actually be to a user.

Limitations

However, it turns out that participants did not generate enough of the right kinds of data to pin down such a function. In order to generate this function, more practice data at more varied difficulty levels would be needed. The author is contemplating ways to coax this data from users for the full study. One such way would be to take control of the difficulty slider away from the participant until we have achieved enough data to construct a reliable curve. However, this would lengthen the time commitment of a participant and likely cause a lower completion rate among respondents.

The sample for this study leaves much to be desired. First of all, the size will need to be increased. We will seek more volunteers from more communities in the future. The IRB has given this study approval to seek volunteers from Reddit, Craigslist and the NDNU campus, so it should be easy to reach more people for the next round of data.

A second problem in the sample is the skew toward individuals who spend significant time playing video games. The most likely reason for this skew was in the presentation of the study - for instance, the name of the website included the words “video game”, and that likely caused non-gamers to believe that they were not qualified to take part. For the next round of data, the messaging will be tweaked to put less emphasis on the “video game” aspect. We hope that this will reduce the skew in the sample.

Since this study assumed that frustration is an integral aspect of fun, it would follow that an individual’s sensitivity to frustration should impact that individual’s preferred difficulty level. Specifically, according to this model those who are prone to frustration would prefer a higher percentage correct on a task. Szalma & Taylor (2011) found that the personality trait of neuroticism correlated with frustration for dealing with automation. Similarly, the author of the current paper would hypothesize that high measures on the neuroticism trait would correlate with preferences for higher percentages correct on a task. Future studies will seek to establish this correlation.

**References**

Amos, S. (2013). “Watch: Gamification goes Mainstream”. Retrieved November 26, 2013 from http://www.huffingtonpost.com/shawn-amos/watch-gamification-goes-m\_b\_3950834.html

Barbe, W. B., & Swassing, R. H. (1979). *Swassing Barbe Modality Index: SBMI*. Zaner-Bloser Incorporated.

Csikszentmihalyi, M. "Enjoyment and the Quality of Life." *Flow: The Psychology of Optimal Experience*. New York: Harper & Row, 1990. 43-70. Print.

Csikszentmihalyi, M., Abuhamdeh, S., Nakamura, J., Flow, A. E., & Dweck, C. S. (2005). *Handbook of competence and motivation.*The Guilford Press, New York, 598-608.

Csikszentmihalyi, M., & LeFevre, J. (1989). Optimal experience in work and leisure. *Journal Of Personality And Social Psychology*,*56*(5), 815-822. doi:10.1037/0022-3514.56.5.815

Davis, M. S. (1971). That’s interesting. *Philosophy of the Social Sciences*, *1*(2), 309-344.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9-15). ACM.

Dollard, J., Miller, N. E., Doob, L. W., Mowrer, O. H., & Sears, R. R. (1939). Frustration and aggression.

Doob, A. N., & Kirshenbaum, H. M. (1973). The effects on arousal of frustration and aggressive films. *Journal of Experimental Social Psychology*, *9*(1), 57-64.

Duggan, M., & Smith, A. (2013). 6% of online adults are reddit users. *Pew Internet & American Life Project*, *3*.

Dunn, R. ( 1993). Learning styles of the multiculturally diverse. *Emergency Librarian*, *20*( 4), 24– 32.

Entertainment Software Association. (2014). *Essential facts about the computer and video game industry*. Retrieved from <http://www.theesa.com/facts/pdfs/ESA_EF_2014.pdf>, 11/23/14.

Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). Does Gamification Work?--A Literature Review of Empirical Studies on Gamification. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on* (pp. 3025-3034). IEEE.

Haworth, John; Stephen Evans (June 1995). "Challenge, skill and positive subjective states in the daily life of a sample of YTS students.". *Journal Of Occupational And Organizational Psychology*. 68(2) (2044-8325): 109–121.

Hokanson, J. E., & Burgess, M. (1964). Effects of physiological arousal level, frustration, and task complexity on performance. *The Journal Of Abnormal And Social Psychology*, *68*(6), 698-702. doi:10.1037/h0049340

Hudson, N. J. (2011). Musical beauty and information compression: Complex to the ear but simple to the mind?. *BMC research notes*, *4*(1), 9.

Ivory, J. D., & Kalyanaraman, S. (2007). The effects of technological advancement and violent content in video games on players’ feelings of presence, involvement, physiological arousal, and aggression. *Journal of Communication*, *57*(3), 532-555.

Jeronimus, B. F., Riese, H., Sanderman, R., & Ormel, J. (2014). Mutual reinforcement between neuroticism and life experiences: A five-wave, 16-year study to test reciprocal causation*. Journal Of Personality And Social Psychology,107*(4), 751-764. doi:10.1037/a0037009

Johnston, C. A., Moreno, J. P., Regas, K., Tyler, C., & Foreyt, J. P. (2012). The Application of the Yerkes–Dodson Law in a Childhood Weight Management Program: Examining Weight Dissatisfaction. *Journal of pediatric psychology*, jss040.

Krätzig, G. P., & Arbuthnott, K. D. (2006). Perceptual learning style and learning proficiency: A test of the hypothesis. *Journal Of Educational Psychology*, *98*(1), 238-246. doi:10.1037/0022-0663.98.1.238

Kundu, R., & Basu, J. (1991). Frustration reactions as predictors of academic achievement and personality dimensions of school children. *Psychological Studies, 36*(3), 146-155.

“New Report from the NPD Group Focuses on Core Gaming and Core Gamers”. (2013) Retrieved November 26, 2013 from <https://www.npd.com/wps/portal/npd/us/news/press-releases/new-report-from-the-npd-group-focuses-on-core-gaming-and-core-gamers/>.

Ogletree, S. M., & Drake, R. (2007). College students’ video game participation and perceptions: Gender differences and implications. *Sex Roles*, *56*(7-8), 537-542.

Prieger, J. E., & Hu, W. M. (2008). The broadband digital divide and the nexus of race, competition, and quality. *Information economics and Policy*, *20*(2), 150-167.

Roberts, N. (2014). The Reliability of Eyewitness Testimony.

Schmidhuber, J. (2007, January). Simple algorithmic principles of discovery, subjective beauty, selective attention, curiosity & creativity. In *Discovery Science* (*pp*. 26-38). Springer Berlin Heidelberg.

Schmidhuber, J. (2009). Simple algorithmic theory of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. *Journal of SICE*, *48*(1).

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *Autonomous Mental Development, IEEE Transactions on*, *2*(3), 230-247.

Silverman, R.S. (2013). “Latest Game Theory: Mixing Work and Play”. Retrieved November 26, 2013 from <http://online.wsj.com/news/articles/SB10001424052970204294504576615371783795248>

Szalma, J. L., & Taylor, G. S. (2011). Individual differences in response to automation: The five factor model of personality. *Journal Of Experimental Psychology: Applied*, *17*(2), 71-96. doi:10.1037/a0024170

Takahasi, D. “Americans spend $23.5B each year on video games”. (2010). Retrieved November 26, 2013 from http://venturebeat.com/2010/05/09/americans-spend-25-3b-each-year-on-video-games/

“Top 10 highest grossing video games ever”. (2012). Retrieved November 26, 2013 from <http://digitalbattle.com/2012/02/21/top-10-highest-grossing-video-games-ever/>

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433-460.

Vygotsky, L.S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Westman, A. S., Alliston, G. R., & Theriault, E. A. (1997). Lack of correlations of sense-modality-oriented indices of learning styles with each other and with classroom tasks. Perceptual & Motor Skills, 84731-737.

Wood, S., Sage, J. R., Shuman, T., & Anagnostaras, S. G. (2014). Psychostimulants and cognition: A continuum of behavioral and cognitive activation. *Pharmacological reviews*, *66*(1), 193-221.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit‐formation. *Journal of comparative neurology and psychology*,*18*(5), 459-482.

Zickuhr, K., & Smith, A. (2013). Home broadband 2013. *Washington, DC: Pew Internet & American Life Project: Pew Research Center*.

**Figure Captions**

*Figure 1:* Flow Channel diagram

*Figure 2:* An approximation of the flow channel (from above, orange = high values, green = low values)

*Figure 3:* An approximation of the flow channel (from a 45 degree view)

*Figure 4:* An approximation of the flow channel (view from the origin into the first quadrant)

*Figure 5:* A slice of the flow channel at *y* = 3

*Figure 6:* A slice of the flow channel at *y* = 5

*Figure 7:* Logistic function with ceiling effect for difficulty values in [0, 1] and floor effect for values in [2, )

*Figure 8:* Possible composition of flow channel slice with inverse floor-ceiling function

*Figure 9:* Possible composition of flow channel slice with inverse floor-ceiling function

*Figure 10:* Possible composition of flow channel slice with inverse floor-ceiling function

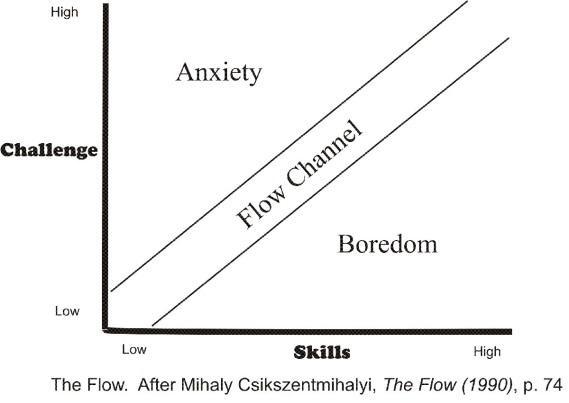
*Figure 11:* Possible composition of flow channel slice with inverse floor-ceiling function

*Figure 12:* Possible composition of flow channel slice with inverse floor-ceiling function

*Figure 13:* Yerkes-Dodson law (retrieved from http://en.wikipedia.org/wiki/File:HebbianYerkesDodson.JPG)

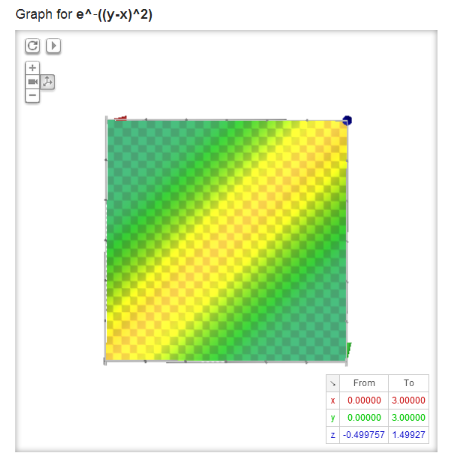
*Figure 14:* Table of percent correct by participant and task type

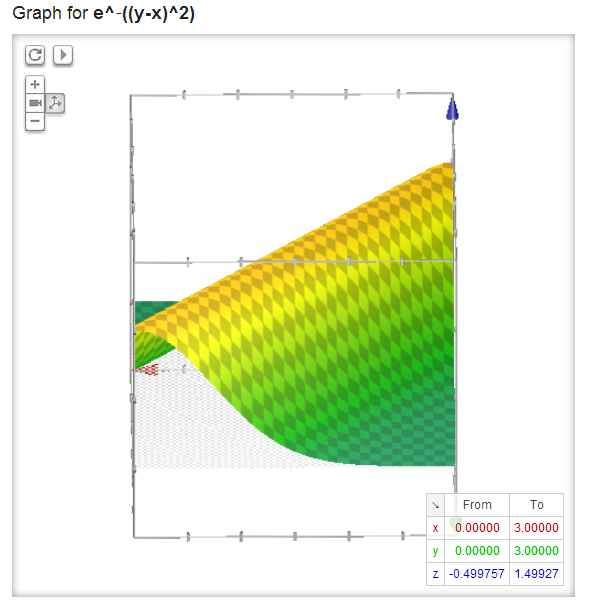
**Figure 1**

Flow Channel diagram

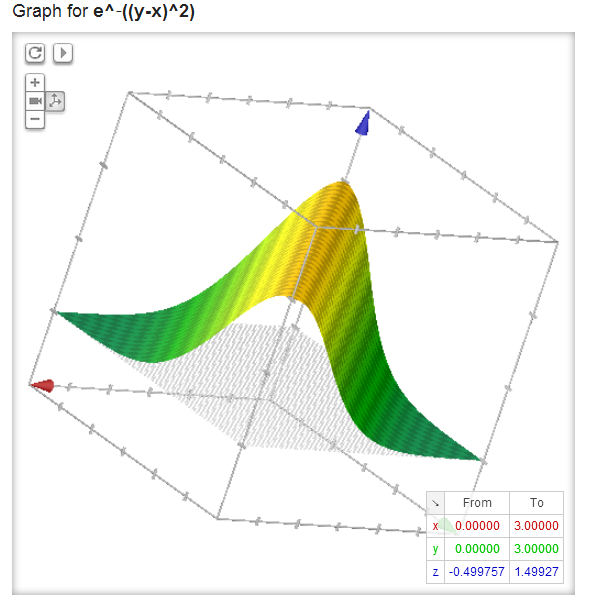
**Figure 2**

**Figure 2**

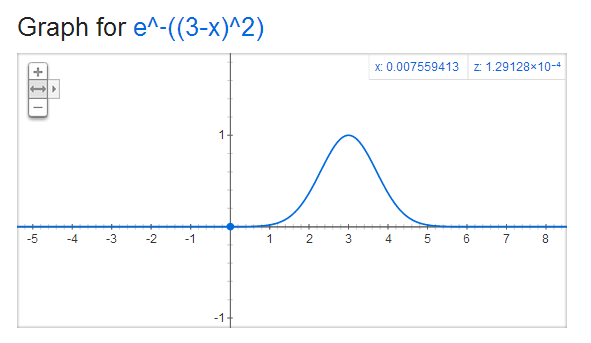
An approximation of the flow channel (from above, orange = high values, green = low values) **Figure 3**

An approximation of the flow channel (from a 45 degree view) 

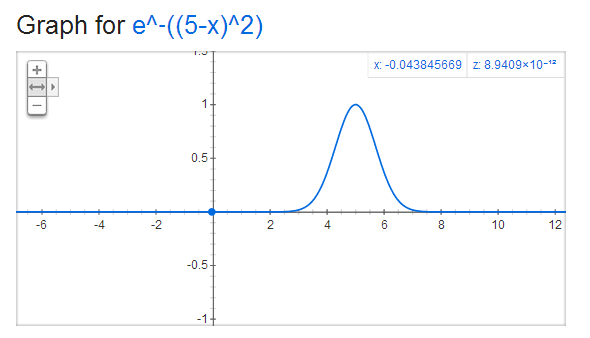
**Figure 4**

An approximation of the flow channel (view from the origin into the first quadrant) 

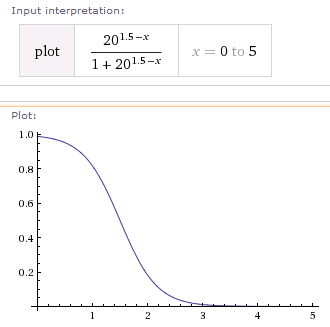
**Figure 5**

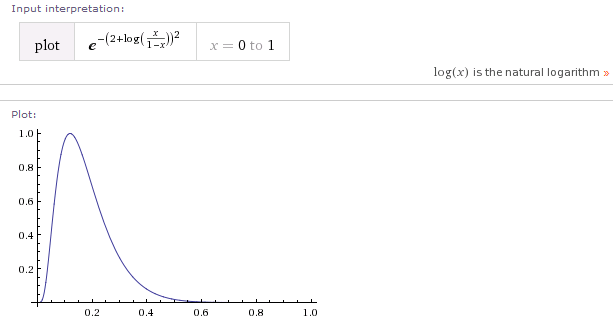
A slice of the flow channel at *y* = 3

**Figure 6**

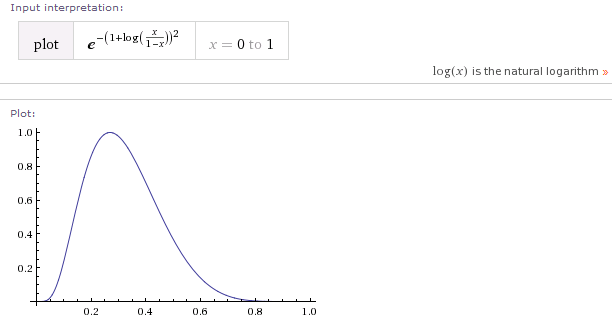
A slice of the flow channel at *y* = 5**Figure 7**

Logistic function with ceiling effect for difficulty values in [0, 1] and floor effect for values in [2, )

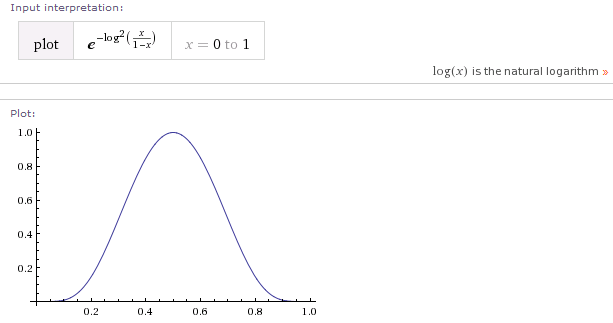


**Figure 8**

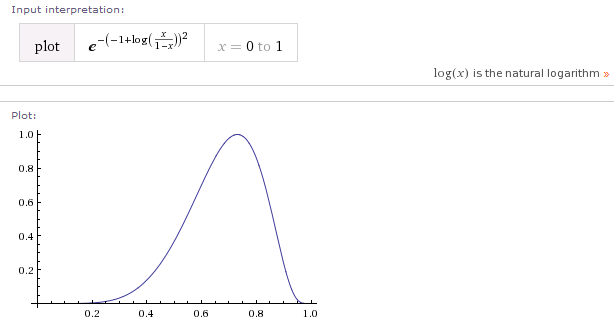
Possible composition of flow channel slice with inverse floor-ceiling function (max at .15)

**Figure 9**

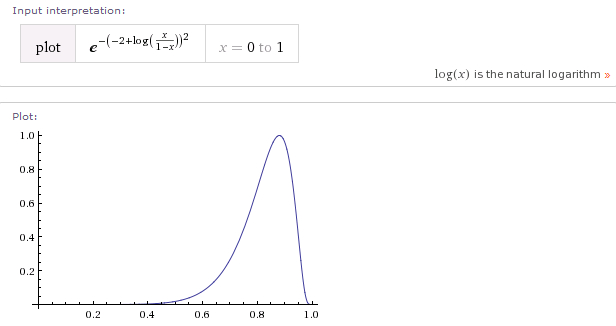
Possible composition of flow channel slice with inverse floor-ceiling function (max at .25)

**Figure 10**

Possible composition of flow channel slice with inverse floor-ceiling function (max at .5)

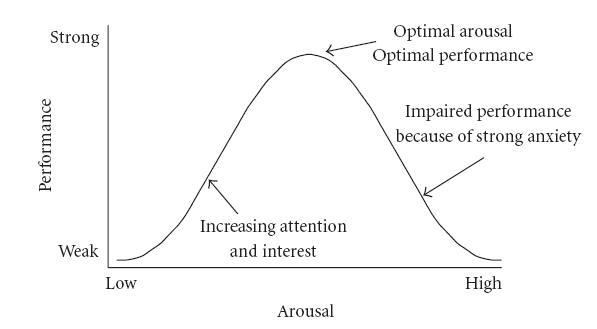
**Figure 11**

Possible composition of flow channel slice with inverse floor-ceiling function (max at .75)

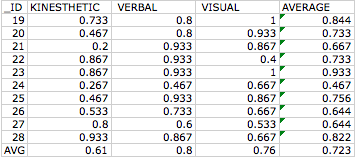
**Figure 12**

Possible composition of flow channel slice with inverse floor-ceiling function (max at .85)

**Figure 13**

Yerkes-Dodson law (retrieved from http://en.wikipedia.org/wiki/File:HebbianYerkesDodson.JPG)

**Figure 14**

Table of average percent correct by participant and task type**Appendix Captions**

*Appendix A*: Visual Task Prompt

*Appendix B*: Visual Task, Answer Revealed

*Appendix C*: Verbal Task Prompt

*Appendix D*: Verbal Task, Answers Revealed

*Appendix E*: Kinesthetic Task Prompt

*Appendix F*: Recruitment Flyer

*Appendix G*: Informed Consent Form

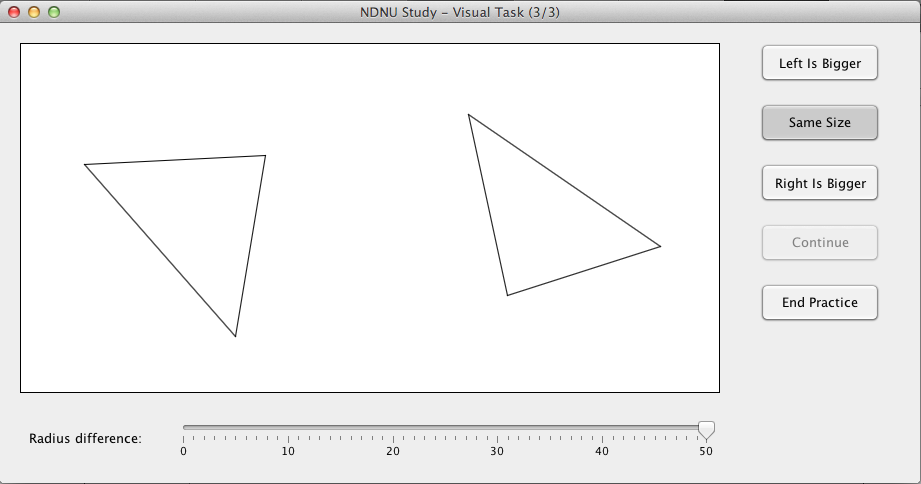
*Appendix H*: Argument from Flow and Range Effects revisited with function estimations

*Appendix I:* Debriefing Form

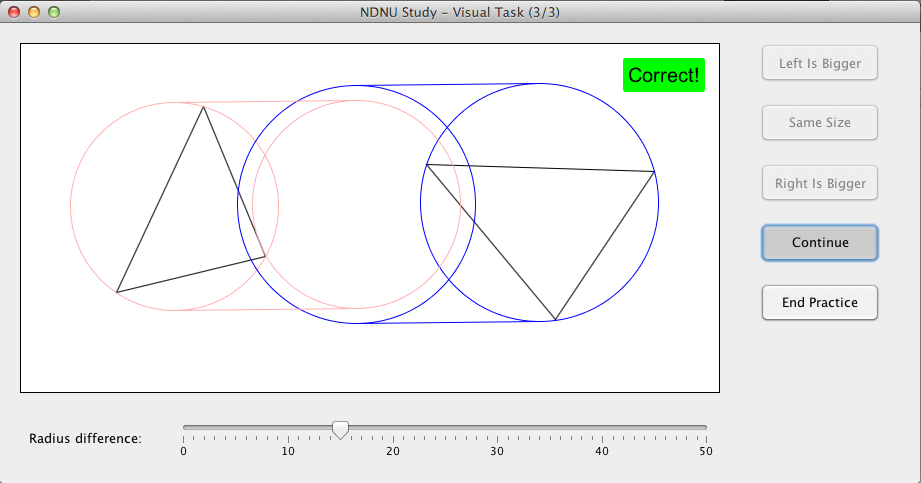
*Appendix J*: Derivation of Composition of Flow Channel Slice with Inverse Floor-Ceiling Function

**Appendix A**

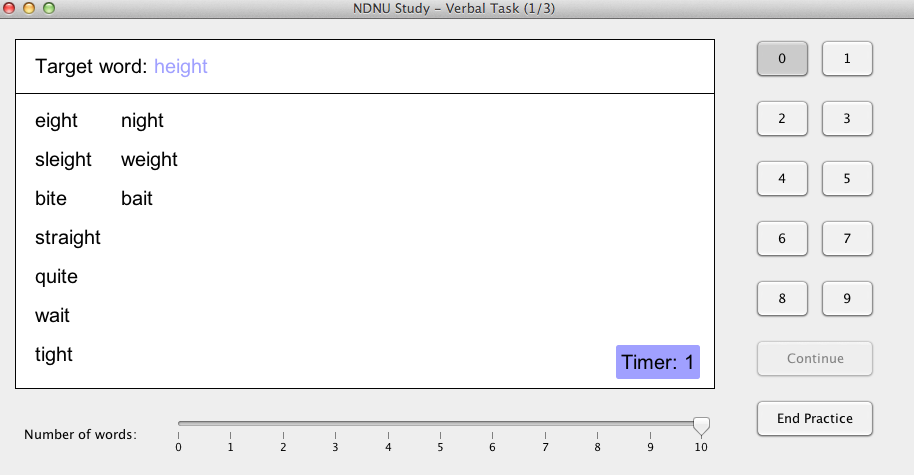
Visual Task Prompt

**Appendix B**

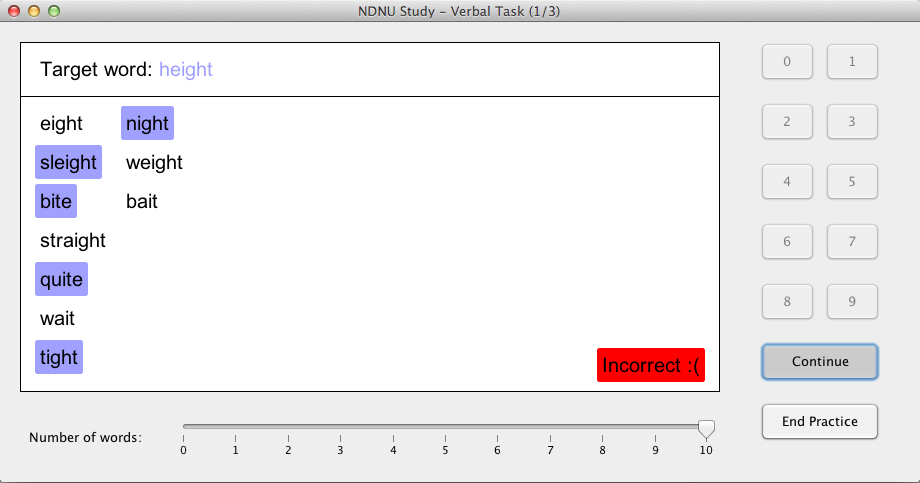
Visual Task, Answer Revealed



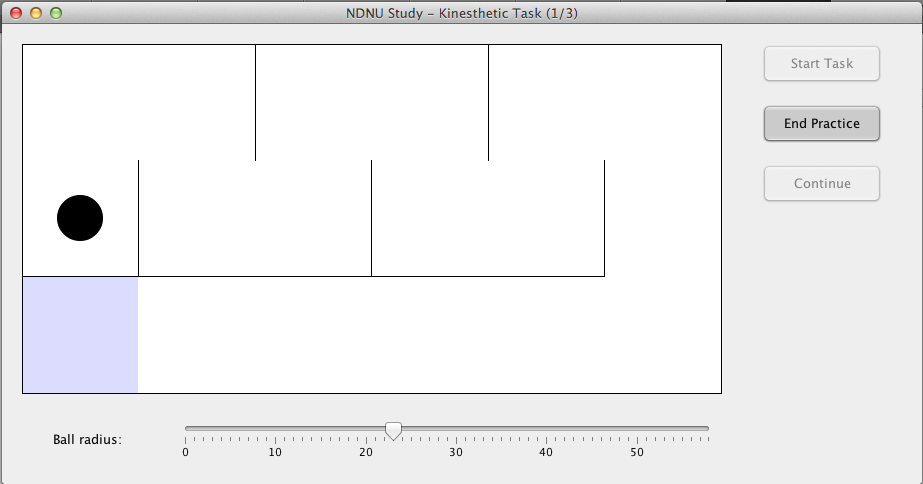
**Appendix C**

Verbal Task Prompt

**Appendix D**

Verbal Task, Answers Revealed

**Appendix E**

Kinesthetic Task Prompt

**Appendix F**

Recruitment Flyer

Seeking adult participants to play video games for a study

No gaming experience necessary

Research is focused on finding the best difficulty level of three different types of games

Participants may be of any gender or race and of

any level of gaming experience

Participants must be 18 years of age or older to participate

Completion of the study involves playing games for 15 to 20 minutes and can be completed at any internet-capable computer with a mouse

To take part in this study, please visit [www.NDNUvideogamestudy.com](http://www.ndnuvideogamestudy.com)**Appendix G**

Notre Dame de Namur Informed Consent Statement

**Preferred Difficulty Levels in Novel Computer-Based Tasks**

You are invited to participate in a research study, the purpose of which is to help understand the factors contributing to preferred difficulty levels in novel computer-based tasks.

INFORMATION

To help us in this task, we will ask you to participate in a study which lasts approximately fifteen (15) to twenty (20) minutes. You will be asked to complete three different types of tasks and to adjust the difficulty level of those tasks to be the most enjoyable for you. There will be three different types of tasks: visual, verbal, and kinesthetic.

For the visual task, you will be asked to mentally draw a circle which goes through all three tips of a triangle. You will be asked to compare the sizes of two of these circles.

For the verbal task, you will be asked to decide how many words in a list rhyme with a target word. You will have ten (10) seconds to complete this task. When you have made your decision, please type in a single number from zero (0) to nine (9) to indicate your choice. For instance, if the target word is “our”, and the list contains “power”, “pour”, “grower” and “sour”, you would answer by pressing ‘2’ on your keyboard (since “power” and “sour” rhyme with “our”, but “pour” and “grower” do not).

For the kinesthetic task, you will be asked to “erase” a line with a virtual “eraser” by moving your mouse in a single stroke. Please do not deviate from the line when doing this.

For each of these types of tasks, you will be able to adjust the difficulty level between trials for the first fifteen (15) trials. You will then be asked to set the difficulty level to a constant value for the next fifteen (15) trials. Please choose carefully when doing so, as once you make your choice this value cannot be altered for the following trials.

A target of 60 subjects has been set for this trial, but that number may vary based upon interest expressed by the population.

RISKS

There are no foreseeable risks associated with this study.

BENEFITS

The results of this study will help benefit scientific understanding of what people want to experience when interacting with computer-based tasks. Such information will be beneficial to

the design of both educational and recreational video games.

CONFIDENTIALITY

The data obtained in this study will be treated as confidential and will be stored securely on a server. You will not be asked to contribute any identifying information. Your responses will be encrypted before being sent to the server, and no personal information about you or your computer will be sent with your responses. The applet with which you are about to interact will not place any cookies on your computer and will not transmit data about the session to any third parties. Data from your responses will be saved for at least three (3) years after completion of the study. Data will be destroyed no more than ten (10) years after completion of the study.

COMPENSATION

You will receive no monetary compensation for participating in this study.

CONTACT

If you have any questions at any time about the study or its procedures, you may contact the principal investigator, Steven Riley, at 619-757-8799 or at [NDNUVideoGameStudy@gmail.com](mailto:NDNUVideoGameStudy@gmail.com). If you feel you have not be treated according to the descriptions in this form, or your rights as a participant have been violated during the course of this study, you may contact the Research Integrity Officer (RIO) at the office of the Provost at NDNU by calling (650) 508-3494.

PARTICIPATION

Your participation in this study is voluntary. You may decline to participate and have your data withdrawn at any time before, during or after completing the tasks in this applet. If you decide to participate, you may discontinue participation at any time without penalty or loss of benefits to which you are otherwise entitled.

CONSENT

If you are not eighteen (18) years of age or above, you are not able to give consent to complete this study. By clicking “I agree”, you state that you have read this consent form and that you understand the above information. Please print out this page for your records.

**Appendix H**

Argument from Flow and Range Effects revisited with function estimations

Csikszentmihalyi gives a visual description of the “flow channel” in a diagram (please see figure 1). The *x*-axis is “skills” and the *y*-axis is “challenge”, and there is a thin region along a 45-degree incline where flow is reached. Above the flow region is an area corresponding to anxiety, and below is a region representing boredom. (Csikszentmihalyi, 1990)

A model for this diagram would be to assign a function *z* = *f*(*x*, *y*) which is maximum in the flow channel and zero in the boredom and anxiety regions. In this function, z is a measure of enjoyment, x is a measure of skills, and y is a measure of difficulty of task. Preferably this would be a simple function that is also smooth. One such function which satisfies these criteria is z = *e*-(*y-x*)^2, which is graphed in figures 2 through 4. In these figures, orange (light colors) represents high values of *z* enjoyment and green (dark colors) represents low values of enjoyment.

At a given time, any individual brings a certain set of skills to a task. In terms of the mathematical model described above, this means that the value of x for that individual is fixed at the time of the task. So, for a fixed value of x, we will get a graph of z (enjoyment) vs. y (difficulty). Such a graph would resemble a bell curve, as shown in figures 5 and 6. The higher the skill, the farther to the right the bell is shifted. This z vs. y curve represents a “slice” through the flow graph of figure 1 along a vertical line.

For the second part of this model, consider a task that has a “floor effect” and a “ceiling effect”. A “Ceiling effect” is the result of a difficulty range where the participant gets nearly all tasks correct. Correspondingly, a “floor effect” is the result of a difficulty range where the participant does no better than random chance. In this model, it is assumed that the floor effect causes percentage correct to plateau at 0%. The results of this model do not depend on this assumption. However, it is easier to demonstrate these results by use of a concrete function.

Consider a task where a difficulty level between zero and one resulted in a ceiling effect, and a level of two or more resulted in a floor effect. In order to model this function with a smooth curve, one option would be to use a logistic function, as shown in figure 7. The y-axis on the graph is percent correct by a participant, and the *x*-axis is difficulty level. This logistic function approaches 100% for difficulty levels below 1 and approaches 0% for difficulty levels above 2.

The third part of this model is to combine the functions we have into a single composite function. There is a composition of these functions that maps percentage correct to enjoyment level. This function is *z* = *f*(*g*-1). Please see Appendix J for a derivation of this function. Figures 8 through 12 show different possibilities for this composite function based on different constants associated with the original functions. Though they differ in shape, all of these functions share a number of properties. They all are defined for inputs between *x* = 0% and *x* = 100%, they all have a single maximum value somewhere between *x* = 0 and *x* = 1, and all contain the points (0%, 0) and (100%, 0). This would imply that the maximum enjoyment happens at a score that is neither 0% nor 100%, and that enjoyment is minimized at these endpoints.

**Appendix I**

Debriefing Form

Thank you for your participation in this research on the preferred difficulty level of computer based tasks**.** Novel computer-based tasks were used for all participants in this study. The goal of the study was two-fold: to gather information on computer-user’s preferred task difficulty, and to attempt to validate an algorithm that predicts difficulty level. It was hypothesized that participants would all prefer approximately the same percentage of correct trials for a task, and that each would set difficulty levels in such a way as to achieve this percentage correct. If you would like to learn more about design of computer based tasks, please see the references listed below.

Current research has found that participants prefer a moderate level of difficulty for tasks. Your participation was important in helping researchers find that level of difficulty and predict the optimal level for future tasks.

Final results will be available from the investigator, Steven Riley, by 12/15/2014. You may contact me at [misterriley@gmail.com](mailto:misterriley@gmail.com) to receive an email copy of the final report. All results will be grouped together; therefore individual results are not available. Your participation, including your name and answers, will remain absolutely confidential, even if the report is published.

If you have any additional questions regarding this research, please contact me at [misterriley@gmail.com](mailto:misterriley@gmail.com).

**Appendix J**

Derivation of Composition of Flow Channel Slice with Inverse Floor Ceiling Function

As noted above, our model for the flow channel function is *z* = *e*(*y* - *x*)^2. A more general version is to allow a shape parameter k as a coefficient on the exponent: *z* = *ek*(*y* - *x*)^2. Slicing the flow channel is equivalent to setting y to a constant c: *z* = *ek*(*c* - *x*)^2

The general form of a logistic function is *y* = 1/(1 + *ea* + *bx*). Solving for the inverse of this function yields

*a* + *bx* = *ln*((1 - *y*)/*y*)

*x* = (*ln*((1 - *y*)/*y*) - *a*)/*b*

Substituting this x into the flow channel slice:

*z* = *ek*(*c* - (*ln*((1 - *y*)/*y*) - *a*)/*b*)^2

Let us define new constants *d* = *k*/*b*2 and *f* = *cb*2. The function simplifies to

*z* = *ed*((*f - a*) - *ln*((1 - *y*)/*y*))^2

Defining *g* = *f - a*, we now have this function in two constants:

*z* = *ed*(*g* - *ln*((1 - *y*)/*y*))^2

Examples of this function are graphed in Figures 8 through 12. Each of these functions has *d* = 1. The values of *g* are integers that range from -2 to 2.