Experience and Intertemporal Choice

by

Evan Fradkin

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Behavioural Science Group

Warwick Business School

University of Warwick

August 2020

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my advisors Daniel Read, Tigran Melkonyan, and Elliot Ludvig for their support throughout the years. Our discussions inspired me and lead me to further engage with my research. I would like to thank John Taylor and Ty Hayes for their help implementing the lab experiments. Also, a special thanks to my friends and colleagues at WBS who participated in the beta testing of my experiments. I would also like to thank Ivo Vlaev, Max Finne, and Mehmet Chakkol for their help designing and implementing the randomised control trials. I would also like to thank Nick Chater for organising his weekly lab meetings; it was an especially useful forum to present some of my early ideas during my studies. Lastly, I would like to thank my family and friends for their immeasurable support.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented in this thesis was carried out by the author with the exception of the data collection detailed in Chapter 5, which was carried out by the module leaders.

Abstract

People need to make decisions over time (intertemporal choice) such as choosing some money now vs more money later or investing time in studying now vs enjoying leisure later. Two factors that can influence these decisions are people's past experiences and feedback they have received on past decisions. However, these two factors are often overlooked when it comes to formally studying decision making, particularly when it comes to decisions over time. This thesis investigates both factors via novel intertemporal choice from experience experiments as well as a randomised control trial testing the effects of different forms of feedback.

The first chapter gives an overview of intertemporal choice and decisions from experience. It begins by discussing what an intertemporal choice is — trading off something now for something later (and vice versa) — and how intertemporal choices are presented in lab experiments as primarily paper-based decisions, particularly when it comes to experiencing delays. Next, an overview of the intertemporal choice models and their assumptions and how these models evolved to capture behaviour is provided. Classical intertemporal choice effects are then presented and discussed. Afterwards, the literature on decisions under risk is introduced, focussing on the decisions from experience paradigm and the description-experience gap. One of the main findings from the decisions from experience literature is that people's decisions diverge depending whether they are using descriptive vs experiential information. This is often times referred to as the description-experience gap. The next section of Chapter 1 covers the various paradigms of intertemporal choice and how experience has been incorporated into the different paradigms to better understand intertemporal choices outside of paper-based decisions. The last section of Chapter 1 reviews experiments that are considered intertemporal choice from experience experiments that are most similar to the experiments conducted in this thesis.

In Chapter 2, two novel intertemporal choice from experience experiments are presented. The vast majority of research on intertemporal choice conducted with

humans has been paper-based, that is, the delays and rewards associated with each alternative are explicitly written. Nor has the vast majority of past research systematically tested how intertemporal choice from experience might differ from intertemporal choice from description. These two studies investigated whether there is a description-experience gap in intertemporal choice. Secondly, they test whether an intertemporal choice effect, the common difference effect, appears in experience. Thirdly, the role of time perception, a potential mediator of the description-experience gap in intertemporal choice, is investigated. Overall, the results from Experiment 1 suggested that there isn't a description-experience gap in intertemporal choice because of the appearance of the common difference effect in both description and experience. However, in Experiment 2, using pairs of delays that were more similar, the common difference effect disappeared in experience. A key reason for the selective appearance of the common difference may be uncertainty around the delays, which is further explored in Chapter 3.

In Chapter 3, the role of uncertainty in a potential description-experience gap in intertemporal choice is tested. Previous intertemporal choice from experience studies that incorporate uncertainty draw different conclusions as to whether there is a description-experience gap in intertemporal choice. The experiment in Chapter 3 attempts to clarify whether uncertainty may be a potential cause of a description-experience gap in intertemporal choice. Moreover, whether uncertainty moderates the size of description-experience gap and the size of the common difference effect in description and experience is examined. Consistent with the previous experiments, we find a common difference effect. When uncertainty around the delays was increased the size of the common difference effect was somewhat diminished relative to when there was less uncertainty around the delays.

In Chapter 4, a randomised control trial on social comparison-based feedback is presented. This randomised control trial provides a more ecological valid test of intertemporal choices than lab based studies. Previous randomised control trials have investigated the role of giving social comparison-based feedback; however, few studies have directly tested the effect of different types of rank-based feedback in non-traditional classroom settings. In this chapter, the influence of rank-based

feedback vs mean-based feedback vs absolute-based feedback on subsequent quiz scores is compared. Overall, different social comparison-based feedback had little effect on quiz scores.

In the final chapter of the thesis, Chapter 5, the findings and contributions of the thesis are reviewed. The main contributions of this thesis show that intertemporal choice effects, specifically the common difference effect, can be replicated on the order of seconds and in experience. Furthermore, this thesis also shows some evidence for a description-experience gap in intertemporal choice and that the gap may be moderated by uncertainty around the delays when people make decisions from experience. Also, this thesis provides a novel paradigm to test whether intertemporal choices from description differ from intertemporal choices from experience. Moreover, this thesis provides a test of intertemporal choice from experience in the real world.

Chapter 1. Literature Review

1.1. Intertemporal Choice

People need to make decisions over time in their everyday lives. These decisions often necessitate trading something in the present for something in the future. For example, these trade-offs could involve forgoing a Maserati and buying a Honda Civic in order to have a higher quality of life in retirement. Trade-offs are not only made in financial decisions; trade-offs in health include avoiding unhealthy foods in the present to enjoy a higher quality of life in the future. The trade-offs made in everyday life are referred to by many different names such as delay discounting, delay of gratification, or temporal discounting. However, the fundamental idea behind each of these terms is that receiving something worse or better has a cost in terms of the relative timing. In this thesis, these choices will be referred to as intertemporal choices.

In order to analyse intertemporal choices, they are categorised into two alternatives: Smaller Sooner (SS) or Larger Later (LL). In the last example above, consuming unhealthy food now would be the SS option and a higher quality of life would be LL option. However, typically, they are presented as money earlier or later choices (MEL) such as £50 now or £100 tomorrow (Cohen *et al.*, 2016). One way to go about understanding how people make intertemporal choices is to define what the normative or "right" intertemporal choice is. The most widely regarded normative model of intertemporal choice is the (exponential) discounted utility model (Samuelson, 1937). The discounted utility model suggests that the utility of an alternative should be the consumption of the reward discounted exponentially with the delay. Below equation 1 gives the discounted utility (denoted as U), that is value or worth, according to the discounted utility model. Equation 1 below shows the exponential discount function of the discounted utility model as proposed by Samuelson where X is the reward, δ is th discount factor, and t is the time horizon (t time periods into the future).

(1)

$$U(\mathrm{X},t) = U(\mathrm{X})\delta^{t}$$

The discounted utility model makes a number of simplifying assumptions. One assumption being that there is a single discount factor that does not change for a given individual over time or across contexts. The consistency of this discount factor implies that people's preferences are consistent over time (Strotz, 1955). Other normative models of intertemporal choice have also been proposed such as the net present value model (Fisher, 1930). In general, the focus of the discounted utility model is the discounted utility for the individual, while the net present model is more generally concerned about the discounted utility for the marketplace.

Later models of intertemporal choice relaxed key assumptions, such as time-consistent preferences, to better capture behaviour. One such model that relaxed these assumptions is the hyperbolic discounting model. The hyperbolic discounting model is the most dominant alternative to the standard model of intertemporal choice. While the hyperbolic discounting model has been formalised in various ways, each formulation attempts to describe behaviour not captured by the discounted utility model (these behaviours are discussed below) (Ainslie, 1975; Mazur, 1987; Green, Fristoe & Myerson, 1994). Equation 2 shows a functional form of the hyperbolic discounting model used by (Mazur, 1987). In Equation 2, X is the reward, k is the discount factor, and *t* is time.

(2)

$$U(\mathrm{X}, t) = \frac{X}{1 + kt}$$

Hyperbolic discounting can describe behavioural phenomenon that exponential discounting cannot. If there are two alternatives closer in the present and two alternatives in the future, people are typically less patient for the pair of alternatives closer to the present and more patient for the pair of alternatives further in the future. This is an example of the common difference effect that will be further

explored below. Hyperbolic discounting captures that people discount alternatives similarly when they are in the future. Whereas when the alternatives are closer to the present, the Larger Later alternative is discounted more heavily than the Smaller Sooner alternative. For example, if a person discounts exponentially and prefers £15 tomorrow to £10 today, this would mean that their daily discount factor (delta) is greater than 0.66. By using a daily discount factor of 0.7 we can find the utility of each alternative for an exponential discounter as shown below:

Alternative 1: $U(10,0) = £10 * .7^0 = 10$
Alternative 2: $U(15,1) = £15 * .7^1 = 10.5$

In the above example, a person with who is an exponential discounter with a discount factor of 0.7 might choose alternative 2, the Larger Later alternative (LL) as it is worth five more utils than alternative 1. That person might also prefer £15 in 15 days to £10 in 14 days as shown below

Alternative 1: $U(10, 14) = £10 * .7^{14} = .06$
Alternative 2: $U(15, 15) = £15 * .7^{15} = .07$

This is an example of time consistency. A crucial component of time consistency and the discounted utility model is that the per period discount factor is constant. What hyperbolic discounting allows for is dynamic inconsistency, that is people choose as if their discount rates change with the length of the delays. In terms of discount rates, this typically means high discount rates in the present and low discount rates in the future (Strotz, 1955). For example, a hyperbolic discounted facing the same decision as an exponential discounter might prefer $10 today to $15 tomorrow as shown below:

Alternative 1: $U(10, 0) = 10 * \frac{1}{1+.7*0} = 10$
Alternative 2: $U(15, 1) = 15 * \frac{1}{1+.7*1} = 5.88$

However, after the alternatives are pushed into the future their preferences may reverse and subsequently, they could prefer alternative 2.

Alternative 1: $U(10,14) = 10 * \frac{1}{1+.7*14} = .9$

Alternative 2: $U(15,15) = 15 * \frac{1}{1+.7*15} = 1$

Another model similar to hyperbolic discounting is quasi-hyperbolic discounting (Phelps & Pollak, 1968; Laibson, 1997; McClure *et al.*, 2004, 2007; Benhabib, Bisin & Schotter, 2010). Quasi-hyperbolic discounting has been proposed to account for people's preferences for immediate rewards, otherwise known as present bias (O'Donoghue & Rabin, 1999a, 1999b, 2015). In essence, quasi-hyperbolic discounting is exponential discounting and an extra weight to capture people's preferences for immediate rewards (Equation 3).

(3)

$$U(X,t) = \begin{cases} X \; if \; t = 0 \\ X\beta\delta^t \; if \; t > 0 \end{cases}$$

The dynamic inconsistency that hyperbolic discounting captures is often characterised as patience, self-control, or lack thereof. The vast majority of experiments in intertemporal choice have found that people's behaviour is closer to hyperbolic than exponential (Frederick *et al.*, 2002). The above examples where people are asked to choose between a Smaller Sooner option (alternative 1) and a Larger Later option (alternative 2) are the canonical intertemporal choice tasks.

1.2. The Common Difference Effect and Time Perception

1.2.1. Common Difference Effect

The common difference effect is when discounting over an interval varies as a function of the onset of an interval. The common difference effect occurs when both delays are pushed further into the future (Loewenstein & Prelec, 1992) as shown in

the examples in the above section. In monetary terms, a person might prefer £10 today vs £15 tomorrow but prefer £15 in 31 days to £10 in 30 days. A common example of the common difference effect in non-monetary terms given by Thaler (1981) is:

(A) **Alternative 1**: One apple today.
Alternative 2: Two apples tomorrow.
(B) **Alternative 1**: One apple in one year.
Alternative 2: Two apples in one year plus one day.

If a person's discount rate was constant deciding between alternative 1 and alternative 2, then choice set A and B would result in identical decisions. However, if one was a hyperbolic discounter, which the literature suggests is most people, one would choose alternative 1 in choice set A and alterative 2 in choice set B. Figure 1 visually shows the common difference effect. In panel A, a person might prefer £50 now to £100 when both rewards are closer in time. Panel B shows that when both rewards are moved further out into the future £100 is preferred at time zero.

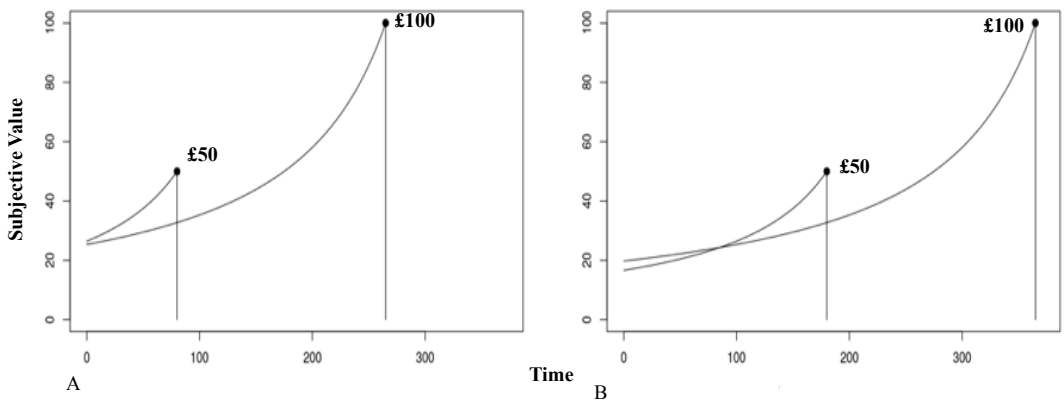

Figure 1: (A) Subjective values for early delays (B) Subjective value for late delays.

1.2.2. Time Perception and Intertemporal Choice

Subjective time perception has also been shown to influence intertemporal choices. In one of the first studies to empirically investigate the effect of time perception on intertemporal choice, Zauberman, Kim, Malkoc, and Bettman (2009)

found that subjective time scaled non-linearly with objective time, which led to hyperbolic like behaviour. They also showed that taking into account subjective time perception, people's behaviour was closer to exponential discounting. Other findings show that subjective time perception can lead to subadditivity, which is when discounting over a time interval is steeper when the interval is divided into sub-intervals than discounting (Read, 2001; Scholten & Read, 2006). Other research has also show non-linear time perception could be reason for the appearance of intertemporal choice effects such as the sign effect (Han & Takahashi, 2012).

A well-known finding in the time perception literature is that people overestimate short delays and underestimate long delays (Stevens & Greenbaum, 1966). Anything that might exaggerate or constrain this principle could have implications for intertemporal choice. For example, imagine that choice set A below shows the true delays and that choice set B shows what might happen if people underestimated both delays in choice set A. Relative to choice set B, people will be more likely to choose alternative 1 in choice set A than in choice set B.

(A) **Alternative 1**: One apple in 5 days.
Alternative 2: Two apples in 25 days.
(B) **Alternative 1**: One apple in 3 days.
Alternative 2: Two apples in 15 days.

This differential under and overestimating may occur under a number of circumstances. To better understand when delays might be estimated differently, we can look to the domain of risk where there is a robust literature that describes when probabilities are weighted differently, depending on whether people learn from about them from description or experience.

1.3. Decisions from Experience

All of the above examples of intertemporal choice include explicitly described delays. This is because intertemporal choice with humans has largely been studied using described delays. That is, the delays are explicitly written out.

However, in everyday life people often have to make intertemporal choices from experience. For example, choosing to wait for the bus or walking instead. Much of the early research in intertemporal choice was conducted using animals which largely precludes the possibility of describing delays (Ainslie, 1975; Mazur, 1987; Rachlin & Green, 1972) because many animals are not able to read. In other domains, how people make decisions from experience has been more thoroughly investigated. In this section I highlight findings in risk that may implications for intertemporal choice.

Many insights from decisions under risk have yet to be imported into research on decisions over time. One particular stream of research that is relevant to intertemporal choice is the description experience gap. There are two primary ways in which people can learn about the world, either through description or experience. For example, if a person wanted to find a path through the forest, they could rely on their past experience of walking through the forest to navigate or they could use a written map of the trail. Throughout the past 10 years, researchers have systematically documented how decisions from description diverge from decisions from experience. The basic motivation for this research is simple. In most judgement and decision making tasks with humans the probabilities and outcomes associated with each alternative are described. However, people make risky decisions every day in situations where the probabilities and/or outcomes are not described. For example, imagine crossing a street that you have crossed many times before. When you decide to cross the street, you have some idea of the probability of crossing the road safely; however, there is not a sign next to the street describing the probability of safely crossing the road. This is the distinction between risk and uncertainty. *Risk* is when the probabilities are known and *uncertainty* is when the probabilities are unknown (Knight, 1921). In a typical task structured to compare decisions from description and experience, people are asked to choose between two alternatives:

Alternative A: £3 with 100% probability.
Alternative B: £4 with 80% probability, otherwise £0.

One of the most dominant model of decisions from description, prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), predicts that people will overweight rare events. A novel feature of prospect theory is its probability weighting function, which captures the non-linear weighting of probabilities.

Conversely, research on decisions from experience has shown that when people have experiential knowledge of the probabilities, they underweight the rare event (Wulff, Mergenthaler-Canseco & Hertwig, 2018). In the choice presented above, the rare event is defined as 20 per cent probability of receiving £0. If people have experiential knowledge of the probabilities associated with A and B, then they will choose B because they underweight the rare event. This is because they underweight the probability of receiving £0. In contrast, the dominant model of decisions by description would predict that people would choose A because they would overweight the 20 per cent probability of receiving £0.

The decisions from experience tasks utilise several different sampling procedures that have important implications for studying decisions from experience, specifically how people received feedback based on their choices. In the *full-feedback paradigm*, people experience the alternative for their selected choice and are also given information about the foregone alternative. In the *partial feedback paradigm*, people also experience their selected alternative. However, they are not given information about the foregone alternative. In both the partial and full-feedback paradigms, every choice is consequential — that is their remuneration is contingent on their choices. In the *sampling paradigm*, participants are able to sample each alternative without consequence. Next, people move into the second phase where they make a one shot consequential decision between the alternatives they experienced in the sampling phase (Hertwig & Erev, 2009).

Understanding the differences between these paradigms for decisions from experience is important because of the variations in sampling and feedback are key reasons why the gap appears. One of the first studies examining the description-experience gap utilised the partial feedback design (Barron & Erev, 2003) and found

that decisions from description diverged from decisions from experience because people underweighted rare events. In another study conducted shortly afterwards, the sampling paradigm was utilised to understand how people make one- shot decisions from experience and they also found that people underweighted rare events (Weber, Shafir & Blais, 2004). Consequently, this left two competing explanations as to why people underweight rare events. Does the gap appear as a result of making repeated decisions, or does it appear because of direct experience with the rare event or lack thereof?

These two questions were addressed in a seminal study by Hertwig, Barron, Weber, & Erev (2004), that employed the sampling paradigm to directly compare description and experience when making one shots decisions. They found a description-experience gap, which suggested that the gap could not simply be attributed to making repeated consequential choices. Rather, the cause of the gap appeared to centre around the rare event. Specifically, they suggested that people underweight the rare event because of small samples and recent samples. People making decisions based on small samples meant that they were unlikely to encounter the rare event thus underestimating its frequency. Furthermore, the last samples participants experienced played an outsized role in their decisions making than did earlier samples (Plonsky *et al*., 2015; Plonsky & Erev, 2017).

Later studies suggested that one of the main reasons for that gap was a sampling error of a statistical nature rather than a psychological one. Fox and Hadar (2006) argue that underweighting in experience as described in previous studies does not occur if the sampled probability distribution is considered rather than objective probability distribution. One study attempted to correct for this sampling error using a matched sampling design (Ungemach *et al*., 2009), which forced people to sample the alternatives equally in the sampling phase unlike the sampling phase used by Hertwig *et al*. (2004) where people could freely sample the alternatives. Despite controlling for sampling error, there was still underweighting of the rare event. A description-experience gap has also been shown to occur even without a rare event (Ludvig & Spetch, 2011).

1.4. Experience and Intertemporal Choice Paradigms

Relative to the domain of risk, the role of experience in intertemporal choice has not been as systematically explored. In the study of intertemporal choice there are two primary paradigms: delay discounting and delay of gratification (Reynolds & Schiffbauer, 2004). Both of these paradigms incorporate experience to varying degrees, although not in ways that allow easy comparisons across paradigms to ascertain the role of experience in intertemporal choice.

A canonical delay discounting paradigm typically asks people description based money earlier or later questions (Thaler, 1981; Cohen *et al.*, 2016) and by doing so one can infer individuals' time preferences by observing their choices between described alternatives. In the delay discounting paradigm, the rewards are typically money. The delay of gratification paradigm is used as a measure of self-control/will power where a person must persist with their decision in order to obtain a reward that is delayed by some period of time. While persisting for the delayed reward a person can always switch to the smaller immediate reward. The first studies to utilise the delay of gratification task involved asking children being placed in a room and being asked to choose between one marshmallow now or two marshmallows later (Mischel & Ebbesen, 1970; Mischel, Shoda & Rodriguez, 1989). Similar to the way patience is coded in the delay discounting paradigm, if a child doesn't wait for the necessary duration to receive two marshmallows, then historically, this was considered a failure of willpower/self-control.

One important distinction between the delay discounting and delay of gratification is how experience is incorporated into each paradigm. In a canonical delay discounting, there is often no experience incorporated into the task. The task is typically description based wherein people choose an SS and LL alternative. However, within the task they often don't experience the delays because they are often on the order of years and months. Moreover, even when the rewards are incentivised, people often do not experience them within the task itself. However, given the nature of the task, the delay of gratification task always includes

experience of the delays within the task. Participants can also consume the rewards during task but whether they consume them immediately is not always clear.

1.4.1. Delay Discounting

To date, a limited number of studies using the delay discounting task have incorporated experience into the delays and rewards. Studies that have included experience in delay discounting paradigms have incorporated experience both via the delays and rewards. In the delay discounting paradigm, these experienced rewards are often referred to as primary rewards. Typical primary rewards are immediately consumable rewards such as food, beverages, or images. When experience is incorporated into the delays, people typically wait for the duration of the delay (Solnick *et al.*, 1980; McClure *et al.*, 2007; Reuben, Sapienza & Zingales, 2010; Crockett *et al.*, 2013). For example, McClure *et al*. (2007) had participants experience both the delays and the rewards. In their experiment, participants fasted before the experiment and then subsequently chose between a Smaller Sooner option or Larger Later option comprised of some smaller amount of liquid now or larger amounts of liquid later.

Other streams of intertemporal choice research interested in understanding Attention Deficit Hyperactivity Disorder and impulsive behaviour have also developed their own paradigms. One paradigm known as the Experiential Discounting Task has participants choose between two alternatives. One of the alternatives is a certain amount of money paired with a delay. The other alternative is comprised of a 35 per cent chance of receiving a titrated amount of money. The amount of money associated with the titrated alternative is adjusted based on the participant's previous choices in order to find an indifference point between the certain alternative and the titrated alternative. During the experiment people actually experience the delays and rewards. In the experiential discounting task, the rewards are coins delivered via a coin dispenser after each choice is made (Reynolds & Schiffbauer, 2004). In a similar task, the Quick Discount Operant Task (Johnson, 2012), participants choose between Smaller Sooner and Larger Later rewards and experienced the delays and rewards. However, the delays were always certain unlike

in the experiential discounting task. Differences between secondary and primary rewards in the delay discounting are consequential for intertemporal choice because people's discount rates have been shown to differ between primary and secondary rewards. When the rewards are directly consumable, people are typically less patient than when the rewards are of a monetary nature (Odum & Rainaud, 2003; Odum, Baumann & Rimington, 2006; Estle *et al.*, 2007).

1.4.2. Delay of Gratification

The delay of gratification paradigm (Mischel & Ebbesen, 1970; Mischel, Shoda & Rodriguez, 1989) does not wrestle with the same methodological issues as delay discounting because the delays and rewards are almost always experienced. Alternatively, in the delay discounting paradigm, the delays and rewards are not experienced. As a result, there are important differences between the delay of gratification and delay discounting paradigms that makes comparing the choices between the two paradigms difficult. Firstly, delay discounting models, such as the exponential and hyperbolic discounting models, cannot be fit to data from the delay of gratification because of the uncertainty of the larger delay. In the delay discounting paradigm, both alternatives have fixed described delays. In a classic delay of gratification task, children are not told exactly when the larger delays, only that it will come later. Moreover, the Smaller Sooner reward is always available immediately (McGuire & Kable, 2012). Accordingly, many studies have asked whether delay of gratification measures "patience" or other constructs. Michaelson, de la Vega, Chatham, and Munakata (2013) find that whether people delay gratification depends on the level of trust via a description-based study. Kidd, Palmeri, and Aslin (2013), using an almost identical setting to the original delay of gratification, place children in either a reliable or unreliable environment. Children in the reliable environment were more likely to delay gratification than children in the unreliable environment. This suggests that children make rational decisions when deciding to delay of gratification. Following this rational approach, McGuire and Kable (2013) show that a person's beliefs about the future should decide whether they delay gratification. If their beliefs about the future delay are Gaussian shaped, e.g., believing a wait is 10 minutes plus or minus a few minutes. However, if their

beliefs about the future delay are heavy tailed e.g., believing initially the delay is short, but might also be very long, then they show people will choose the immediate reward. In a later study (McGuire & Kable, 2015) place people into an environment where it pays to persist and another environment where it does not pay to persist. They find that after learning about each environment, people are able to rationally calibrate their persistence depending on whether they were placed in an environment where it paid to persist or did not pay to persist.

1.5. Similarities Between Risk and Time

Many parallels have been drawn between risk and time (Prelec & Loewenstein, 1991). A commonality particularly relevant to this thesis is the parallel between the common difference effect and the common ratio effect (Allais, 1953). The common difference effect occurs when a constant delay is added to both alternatives. As a consequence, people are more patient for the alternatives with the added constant delay. Similarly, the common ratio effect occurs when the probabilities associated with each alternative are reduced by a common ratio. This reduction shifts people's preferences from the alternative with the smaller reward with higher probability to the alternative with the larger reward with lower probability. For example, when choosing between alternatives in choice set A people often prefer Alternative 1 to Alternative 2. However, when the probabilities are reduced by a common factor of four, people's preferences switch to Alternative 2. With both the common difference and common ratio effect people's preferences switch to the alternative with the larger reward.

(A) **Alternative 1**: £3000 with 100% probability.
Alternative 2: £4000 with 80% probability, otherwise £0.
(B) **Alternative 1**: £3000 with 25% probability, otherwise £0.
Alternative 2: £4000 with 20% probability, otherwise £0.

Previous research has also investigated whether the common ratio effect can also be found in experience (Harman & Gonzalez, 2015). Overall, they found that common ratio effect disappears in experience. This is consistent with other findings that show

effects typically found in description often disappear or are diminished in experience (Dutt *et al*., 2014; Gonzalez, 2013). Given the similarities between risk and time, there are ample reasons to suggest that certain intertemporal choice effects, such as the common difference effect, may also disappear in experience.

1.6. Intertemporal Choice from Experience

Thus far, very few intertemporal choice studies have people learn about the delays and rewards in the same way that people learn about probabilities and rewards from experience. In one study, Dai, Pachur, Pleskac, and Hertwig (2019) had people choose between a sure timing option and a timing lottery. For the sure timing option, participants receive x reward at time t with 100 per cent probability. For timing lotteries, participants receive x reward at time $t1$ with probability p or at time $t2$ with probability 1-p.

In the description condition, people learned about the reward, possible delays, and the probability of each delay via explicit description. In the experience condition, participants learned about the rewards, delays, and probabilities of each delay in a sampling phase. In the description condition, people chose the timing lottery more frequently than in experience when the delay was shorter and rare. When the delay for the timing lottery was longer, the pattern was reversed. This pattern reversal is considered evidence of a description-experience gap. One important thing to note is that unlike other experiential paradigms, people did not actually experience the delays or rewards when sampling. In the experience condition, people were only shown the descriptions after clicking on one of the options and paid a flat participation fee. They do not wait for the duration of the delay or receive the rewards associated with their choices.

Other studies have used different variations of description and experience. Lukinova, Wang, Lehrer, and Erlich (2019) use verbal descriptions (description) and non-verbal representations (experience) of the delays and rewards and had people choose between Smaller Sooner versus Larger Later rewards. For the verbal trials, the delays and rewards were explicitly described. For the non-verbal trials, the

delays were mapped to a modulation rate where the slower the modulation rate, the longer the delay. For the rewards, the larger reward was, the higher the frequency of the sound. Overall, discount factors for short delays between verbal and non-verbal trials were highly correlated ($r = .79$). However, there was only $r = .40$ between verbal and non-verbal tasks when the delays were long. In light of these results, Lukinova, Wang, Lehrer, and Erlich (2019) suggest that animal studies, where no descriptions are used because animals cannot read, are comparable to human studies using short delays. Moreover, these results might scale to human decisions over short time spans but not longer time spans, which may be another source of a potential description-experience gap in intertemporal choice. One important methodological point to note is that time scales between the verbal and non-verbal task were not always the same. For the non-verbal tasks, delays could range into days rather than seconds found in the verbal tasks.

1.7. Interim Summary

In this section, I described how intertemporal choice models evolved to capture people's preferences for more immediate rewards. I also discussed important findings from the decisions from experience literature, specifically the description-experience gap. Moreover, I also reviewed similarities between risk and time and how effects found in description are often diminished in experience. Consequently, I suggested that because of similarities between risk and time a description-experience gap also may exist in intertemporal choice. Next, I covered how the various intertemporal choice paradigms incorporate experience. Notably, very few, studies on intertemporal choice offer a direct comparison of intertemporal choices from description vs experience. The following chapter introduces an experiment that directly compares intertemporal choices from description vs experience similar to the way decisions from experience vs description are compared in the domain of risk.

Chapter 2. Common Difference Effect Gap

2.1. Experiment 1

2.1.1. Introduction

People often need to make decisions about delayed outcomes and sometimes have to rely on their experiences from the past to estimate how long delays may be. One intertemporal choice effect, the common difference effect, is well documented in description, but not in experience. In a canonical intertemporal choice experiment, participants are asked to choose between different described outcomes such as 10 dollars in 5 days vs 15 dollars in 25 days (see Cohen, Ericson, Laibson, & White, 2016 for review). Typically, both the delays and rewards in these experiments are explicitly described; people do not need to learn about the delays and rewards through their own experience. Consequently, whether common difference effect also occurs in experience is unknown.

Thus far, relatively few studies have systematically incorporated experience into delay discounting with humans. Of those that have, the major methods have used either primary rewards, such as chocolate, juice, and other beverages (Brown *et al*., 2009; Jimura *et al*., 2009; McClure *et al*., 2007; Read & van Leeuwen, 1998; Reuben *et al*., 2010) or unpleasant experiences. In general, these studies show that people exhibit more impatience for primary rewards than secondary rewards. Common primary rewards are often food/juice, while secondary rewards are typically money.

In the domain of risk, the role of experience has been more thoroughly investigated in both humans and non-humans (Heilbronner & Hayden, 2015). When knowledge of the probabilities and rewards is gained through experience rather than description (Hertwig *et al.*, 2004; Hertwig & Erev, 2009), there is a divergence in choice, especially when rare events are involved (for a recent meta-analysis; see Wulff, Mergenthaler-Canseco, & Hertwig, 2018). In decisions from experience, rare events are underweighted, whereas in decisions from description, rare events are overweighted (Kahneman & Tversky, 1979). This difference between description

and experience is often referred to as the description-experience gap. The gap has also been shown to persist even without rare events, such as when extreme outcomes are overweighted in experience (Ludvig & Spetch, 2011; Ludvig *et al.*, 2018).

Several reasons have been suggested as to the underlying cause of the description-experience gap, including a reliance on small samples and a reliance on more recent samples (Hertwig & Erev, 2009). A meta-analysis containing 70,000 choices made by over 6,000 participants found that the largest determinant of the gap was indeed a reliance on small samples with recent samples being weighted more heavily (Wulff, Mergenthaler-Canseco & Hertwig, 2018). The description experience gap in risky choice is highly relevant to intertemporal choice because previous studies have suggested there are commonalities between intertemporal choice and risky choice (Loewenstein & Prelec, 1992; Luckman *et al*., 2018; Prelec & Loewenstein, 1991).

One specific area where past research has focused on a link between risk and time is the common ratio effect (Allais, 1953) and the common difference effect (Prelec & Loewenstein, 1991). Harman and Gonzalez (2015) tested whether the common ratio effect also appears in experience and found that in experience, it disappears. While the overall number of reversals in experience did not differ from description, they found that a majority of participants making decisions from experience were consistent and that consistent participants tended to explore more. Accordingly, one should anticipate that at the minimum, intertemporal choice effects should be reduced in experience, particularly, the common difference effect.

Thus far, few intertemporal choice studies have directly asked whether there is a description experience gap (Dai *et al*., 2019; Lukinova *et al*., 2019). One such study claims to have found a description experience gap similar to the gap found in risk. In their experiment, people chose between a sure-timing option, which had an associated sure delay and a timing option, which was associated with multiple probabilistic delays. In the description condition, the delays were explicitly stated (temporal uncertainty). In the experience condition, people learned about the delays via a sampling paradigm (temporal ambiguity). In the description condition, timing

lotteries with shorter rare delays were chosen more often when pitted against a sure-timing option than in experience. As the rare delay grew in length, this pattern was reversed. (Dai *et al.*, 2019). Another study addressing whether the gap exists in intertemporal choice found no gap between description and experience. They used verbal descriptions as description and non-verbal representations (sound) as experience. For the experience trials they paired larger delays with a high modulation rate and smaller delays with a lower modulation rate; and using this method, they found a strong correlation between discount factors in description and experience (Lukinova *et al.*, 2019).

In the present intertemporal choice experiment, participants learn about delays by clicking on the options shown on the screen as coloured buttons and experienced the corresponding delays. Participants then select between either a classical or pop playlist. Their preferred playlist is then paired with the larger delay in the decisions phase. Afterwards participants are offered a series of choices comprised of the delays in the sampling phase where their preferred playlist is paired with the larger delay and their non-preferred playlist paired with the smaller delay.

This experiment improves upon methodology used in past experiments in several areas. Firstly, in this experiment, in the experience condition, participants actually experience the delays in the sampling phase. The delays in the sampling phase are also the same delays used in the decision phase. Other sampling paradigms mix description with experience by simply having people sample descriptions (Dai *et al.*, 2019), participants do not actually experience the delays. Moreover, in our experiment the same delays are also used across description and experience to ensure participants are making equivalent choices unlike in previous experiments where different delays were used in experience than in description (Lukinova *et al.*, 2019).

In the present experiment we use songs as rewards for several reasons. Firstly, we aim to make our results more directly comparable to previous studies using primary rewards. Moreover, we want participant to make direct trade-offs between doing something more pleasurable (listening to music) vs doing something less pleasurable (waiting). Lastly, we wanted choices similar to those that people

make in real life, for example listening to an advertisement in order to consume media such as videos or songs.

A potential mechanism for any observed description-experience gap in intertemporal choice could be due to a difference in time perception between description and experience. Time perception is known to play a role in intertemporal choice (Takahashi, 2005, 2006; Brocas, Carrillo & Tarrasó, 2018). Relying on experience could exaggerate principles of time estimation such as the overestimation of short delays and underestimation of long delays (Stevens & Greenbaum, 1966). If people further overestimate short delays and further underestimate long delays in experience relative to description this could lead to different choice behaviour and consequently a description-experience gap. Taken together, a description-experience gap in intertemporal choice may manifest itself in one of two ways – A general difference in overall patience between description and experience. Alternatively, a difference in the size of the common difference effect between description and experience. A likely mechanism for the description-experience gap, particularly for the latter is non-linear time estimation as detailed above.

In summary we ask the following questions:

1) Do people differ in patience in Description vs Experience?
2) Does a Common Difference Effect appear (a) in Experience and (b) with short delays?
3) Is there a CDE gap?
4) Does time perception relate to impatience?

2.1.2. Methods

2.1.2.1. Participants

In accordance with our pre-registered sampling plan, we collected a sample of 192 participants with 96 participants in the experience condition and 96 participants in the description condition. The sample size is based on a power

analysis with a medium effect size (d = 0.5), alpha = .01 to account for multiple comparisons, and power = .8. Sample size was calculated using G*Power 3.1.

Eight participants in experience and three in description failed the manipulation check and were excluded from the analysis. This left a total of 88 participants in experience and 95 participants in description. Participants were paid a £3 show-up fee and offered up to two pounds for performance. Participants were recruited via Sona at the University of Warwick, and ethics approval was secured through the university's human ethics committee. People arrived in the lab and were provided earphones and seated in front of individual computers. They were asked to place their phones and watches out of sight.

In Experiment 1, we had people pick between different delays paired with musical rewards. They were either told what the delays were (i.e., description) or had to learn about them through sampling (i.e., experience). Therefore, we use rewards that are paired with varying delays on the order of seconds to a couple of minutes. We construct two different types of choice pairs, an early choice pair and a later choice pair. In the early choice pair, the delays are shorter and in the later choice pairs the delays are longer (Table 1).

Groups	**Delays (in S)**	
	Early	Later
Description	5 vs 25	45 vs 65
Experience	5 vs 25	45 vs 65

Table 1: Trial Types by Condition for Experiment 1

2.1.2.2. Design and Procedure

The current study used a 2x2 mixed design shown in Table 2. The between-subject factor was how knowledge of the delays was acquired with two levels: description or experience. The within-subject factor was delay length with two levels: early and late. There were two pairs of delays: two early (5 and 25 seconds) and two late (45

and 65 seconds) (Figure 2). Ethics approval was given by *WBS NER committee approval #001/15-16.*

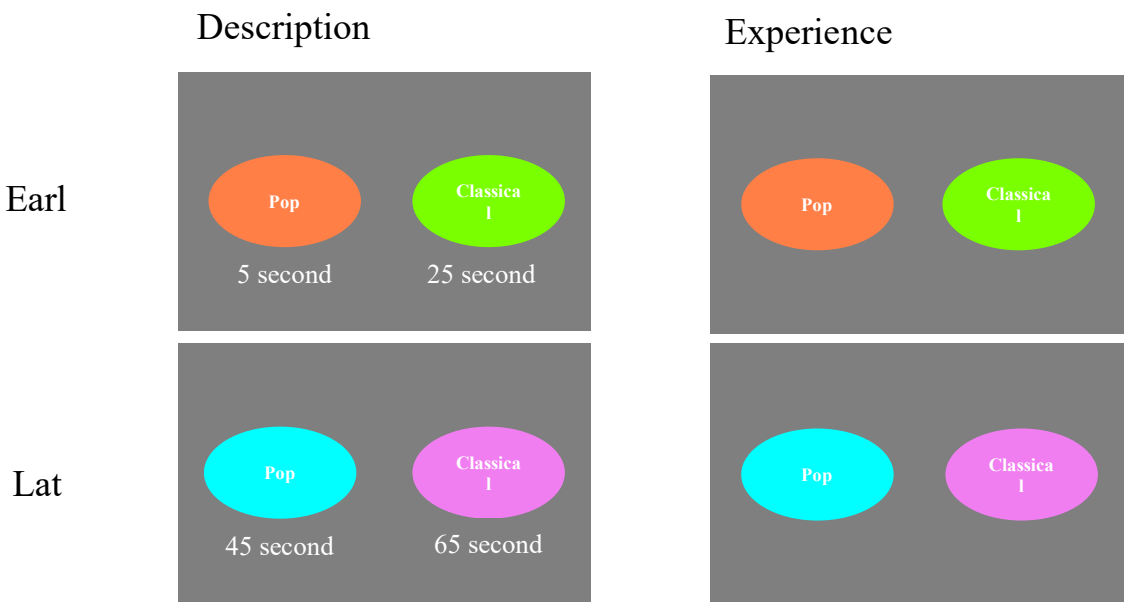

Figure 2: Images of trial types in experiment 1

Phase 1: Sampling. In both groups, participants learned about delays by clicking on the buttons displayed on the screen as coloured ovals and experienced the corresponding delays. In the experience condition, the buttons and delays were the same as the ones in the upcoming decision phase. Each button was sampled four times in a random order followed each time by the appropriate delay. In the description group, a sampling phase was also included, to prevent a differential front-end delay, whereby the experience group had to wait longer to make their choices due to the sampling phase. The sampling phase for the description group differed from the one in experience as all the buttons were grey, and the delays (19s, 38s, 46s, 37s) were not the same as the delays (5s, 25s, 45s, 65s) in the subsequent decision phase. The net waiting time, however, in the description sampling phase was exactly equal to the net waiting time in the experience sampling phase. For each participant, the mapping of the buttons to delays was randomised.

Phase 2: Reward Selection. Participants then viewed two playlists, one that was made up of classical songs and the other made up of pop songs. They then chose

which playlist they preferred to randomly hear a song from. This preferred reward was then paired with the larger delay.

Phase 3: Decisions. Participants were then offered a series of decisions between rewards (music clips from one of the two song genres) at different delays. Options were displayed as coloured buttons with the potential music reward inscribed on them. For the experience group, no further information was provided, but the coloured buttons matched those from the sampling phase. For the described group, the delay for each option was displayed underneath the button. Participants made a total of 16 decisions. Twelve of the 16 decisions were decision trials, which pitted a Smaller Sooner reward against a Larger Later one. The trials were either early trials, where the sooner reward came in 5 seconds and the later reward in 25 seconds, or late trials, where the sooner reward came in 45 seconds and the later reward in 65 seconds. The order of the trials was randomised. Participants were shown every pair six times, meaning that there were 12 decision trials in total. For each participant, their preferred playlist was paired with the larger delay and non-preferred playlist paired with the smaller delay. The sides for the smaller delays and for the larger delays were randomised. In addition to the decision trials, there were four catch trials, which served as a manipulation check and pre-registered exclusion criterion. In the four catch trials, the delays were the same, but the rewards were different. If a participant chose the button with the larger delay on more than one catch trial, they were excluded from the analysis[1].

Phase 4: Time Estimation. In this phase, participants made two separate duration estimates, for a total of 8 time estimates, of the delays they decided between in the decision phase. To record each time estimate, participants in both conditions were randomly shown each delay that appeared in the decision phase. A coloured button, representing the delays, appeared on the screen one at a time in the experience condition with no description of the delay. For the description group, a coloured button appeared on the screen with a description of the delay underneath. After being shown the coloured button, participants pressed the space bar to start the

[1] In the pre-registration, the catch trial was originally equal rewards and different delays.

timer and then the space bar again to stop the timer. For each delay, if a participant's estimate was off by 20 per cent or less, they were rewarded 25 pence, with the possibility of earning up to £2.00. This was repeated twice for each of the four delays.

Phase 5: Musical Reward. After the time estimation task was complete, one of the participant's choices from Phase 3 (Decisions) was randomly selected. Participants then waited the corresponding delay and then experienced their chosen musical reward.

2.1.2.3. Data Analysis

Patience was calculated separately for each condition as the proportion of Larger Later (LL) choices made on the decision trials. LL is defined here as the alternative with the larger delay and preferred reward. Patience was also calculated separately for decisions between early and late options in order to measure the size of the common difference effect. A pre-registered 2x2 repeated measures analysis of variance (ANOVA) with condition (description, experience) and delay length (early, late) tested whether there was differential patience between conditions and early and late trials. All degrees of freedom for one-way ANOVA tests were corrected using the Greenhouse-Geisser correction. Our predefined level of significance was defined at $\alpha = .01$. All hypotheses, methods, and analyses were pre-registered at https://osf.io/yjrc4/. Any exploratory analysis is marked as such in the analysis section.

2.1.3. Results

2.1.3.1. Description-Experience Gap and Common Difference Effect

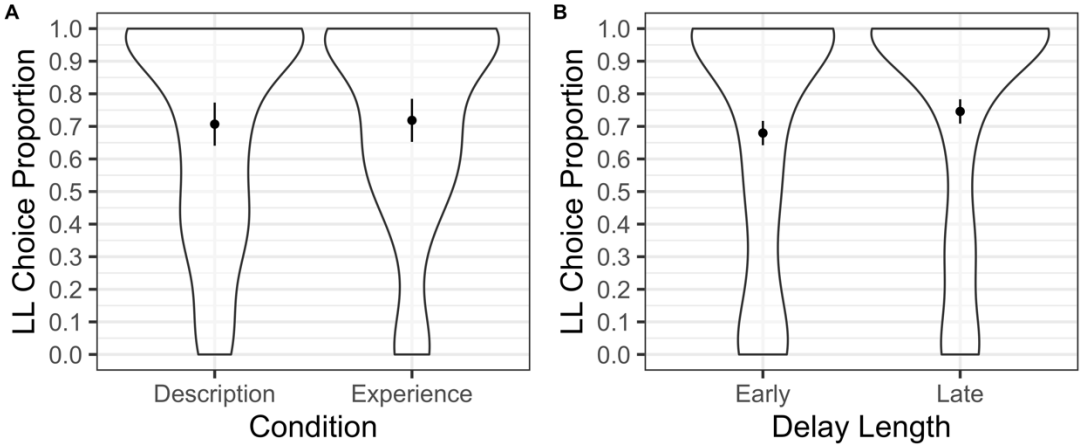

Figure 3: Proportion of LL choices on decisions trials between the description condition and experience condition. The dots show the mean choice proportion while the bars indicate 95 per cent confidence intervals.

Our first research question was whether people would differ in their patience levels between description and experience. Overall, people were not significantly more likely to select the LL option in experience than in description $F(1,181) = 0.06$, $p = .80$, $\eta_p^2 = .0003$ (Figure 3A). Our second research question was whether a common difference effect appears (a) in experience and (b) with short delays? In experience people were significantly more likely to select the LL option with the later delays $t(87) = -1.54$, $p = 0.06$, $d = .16$ (Figure 4). Our third research question was whether there was a difference in the size of the common difference experience between description and experience. People did not choose differently between early and late delays between experience and description $F(1,181) = .02$, $p = .89$ $\eta_p^2 = .0001$ (Figure 4).

Overall people were significantly more likely to take the LL option in the late vs the early condition $F(1,181) = 6.09$, $p = .01$, $\eta_p^2 = .03$ (Figure 3B), although the level of significance was just above our pre-defined threshold. In description, people

were significantly more likely to select the LL option with the later delays $t(94) = -1.97$, $p = 0.03$, $d = .20$.

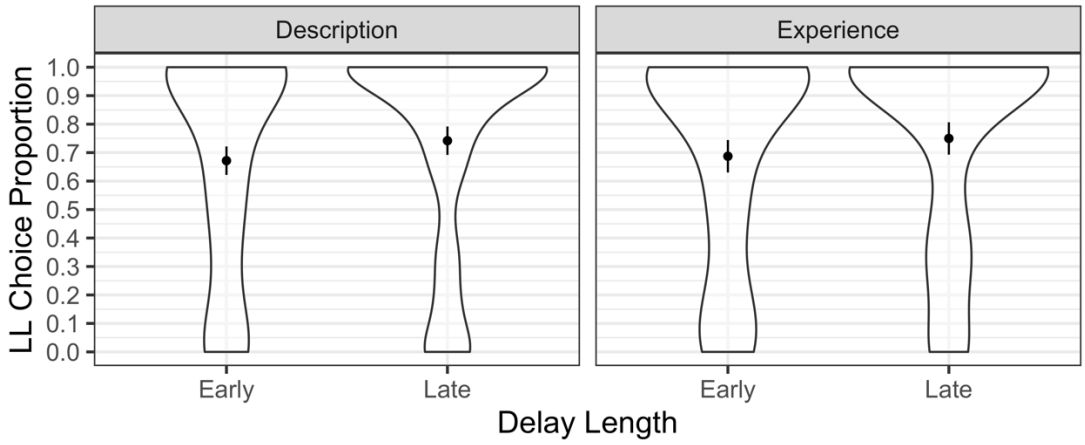

Figure 4: Proportion of LL choices on early and late trials by condition. The dots show the mean choice proportion while the bars indicate 95 per cent confidence intervals within participants.

2.1.3.2. Time Perception

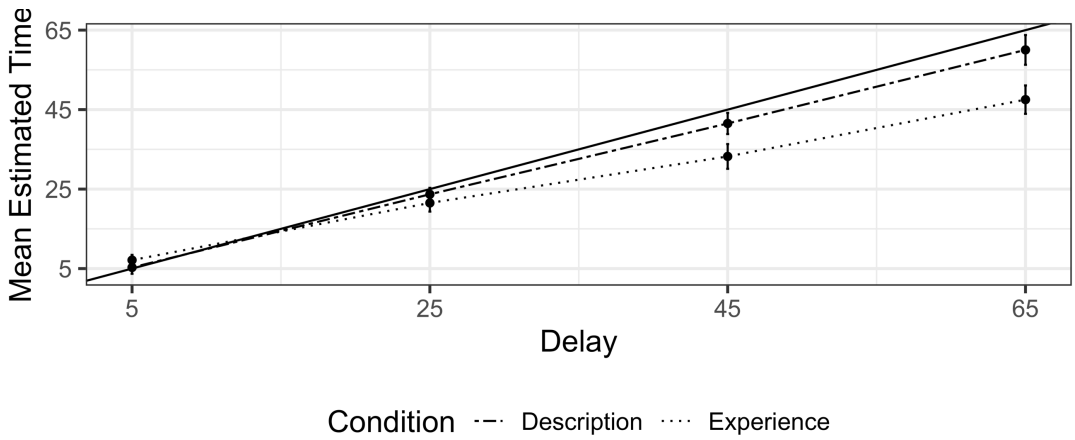

Figure 5: Mean times for each delay by description and experience. The dots show the mean estimated time while the bars indicate 95 per cent confidence intervals within participants. The dotted line shows the experience condition while the dashed line shows the description condition. The solid indicates perfect accuracy.

Figure 5 shows that people in the experience condition underestimated the later delays to a much larger extent than in the description condition. However, the time estimates for the early delays were more similar for description and experience.

Greater underestimation of the later delays could potentially be one of the reasons why we observed a diminished common difference effect in experience.

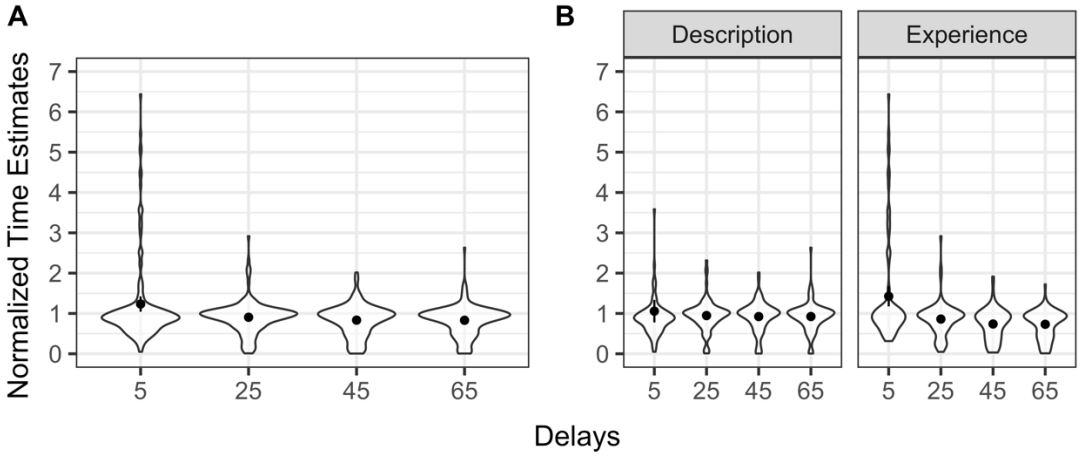

Figure 6: Mean normalised time estimates for each delay. The mean normalised time estimate is the mean of two time estimates divided by the true delay per participant. (A) The mean normalised time estimates collapsed across conditions. (B) The mean normalised time estimates for each delay by condition. A mean normalised time estimate of one would indicate perfect accuracy.

Figure 6 shows the mean normalised time estimates for each delay (Figure 6A) and condition (Figure 6B). Overall, Figure 6 suggests that there were differences in the accuracy of participants' time estimate between delays. This observation was further confirmed by a pre-registered one-way repeated-measures ANOVA with mean normalised time estimates (mean of two estimates divided by the true delay) as the dependent variable and delays (5,25,45,65) as the independent variables $F(1.14,207.25) = 11.15$, $p = .0006$ $\eta_p^2 = .06$.

Figure 6 also suggests that there are differences in the accuracy of participants estimates across conditions. This is supported by an exploratory ANOVA with delays (5,25,45,65) and condition (description, experience) $F(1.14,206.61) = 5.50$, $p = .02$, $\eta_p^2 = .03$, which confirmed a mild interaction between delays and condition. Exploratory *t* tests revealed that there was a significant difference in the mean normalised estimates for the 45 seconds delay between description and experience $t(181) = 3.33$, $p = 0.001$ and significant

difference in the mean normalised time estimates for the 65-s delay $t(181) = 3.67$, $p < 0.001$. All other comparisons across conditions were not significant.

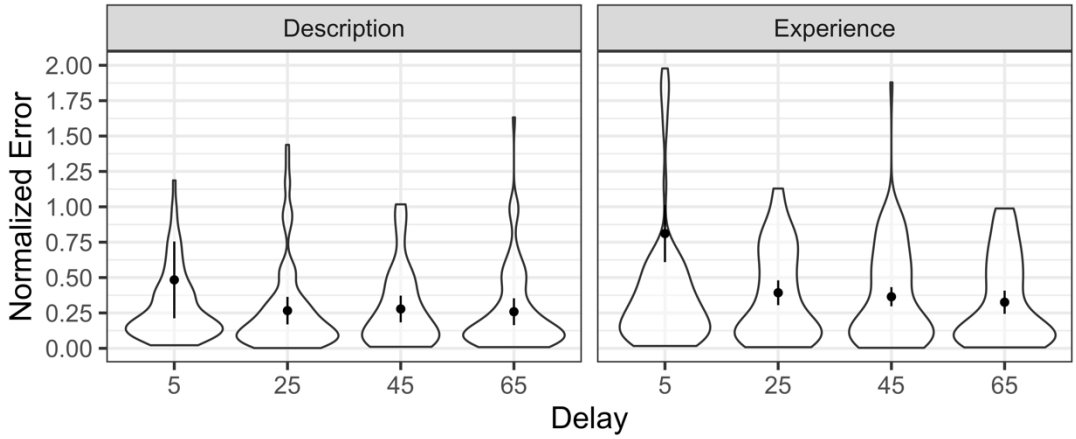

Figure 7: Normalised error for each delay across description and experience. The normalised error is the delay subtracted by the time estimate divided by the delay. The dots show the mean normalised error while the bars indicate 95 per cent confidence intervals within participants.

Figure 7 shows the normalised error for each delay between description and experience. The normalised error is the delay subtracted by the time estimate divided by the delay. In description the normalised error sits closer to 25 per cent, while in experience the normalised error is closer to 40 per cent. The exception in both conditions being the five second delay. Overall, participants' time estimates appear to be scale invariant (Gallistel & Gibbon, 2000).

2.1.3.3. Choice and Time Perception

For our fourth question, we asked whether time perception relates to intertemporal choice. Figure 8A plots the frequency of LL chosen among early pairs against the mean normalised time estimate of early pairs. For early pairs, there was no correlation between patience and time estimates, $r(181) = 0.05$, $p = 0.50$. Figure 8B plots the frequency of LL chosen among later pairs against the mean normalised time estimate of later pairs. For later pairs, there was also no correlation between patience and time estimates, $r(181) = 0.11$, $p = 0.15$.

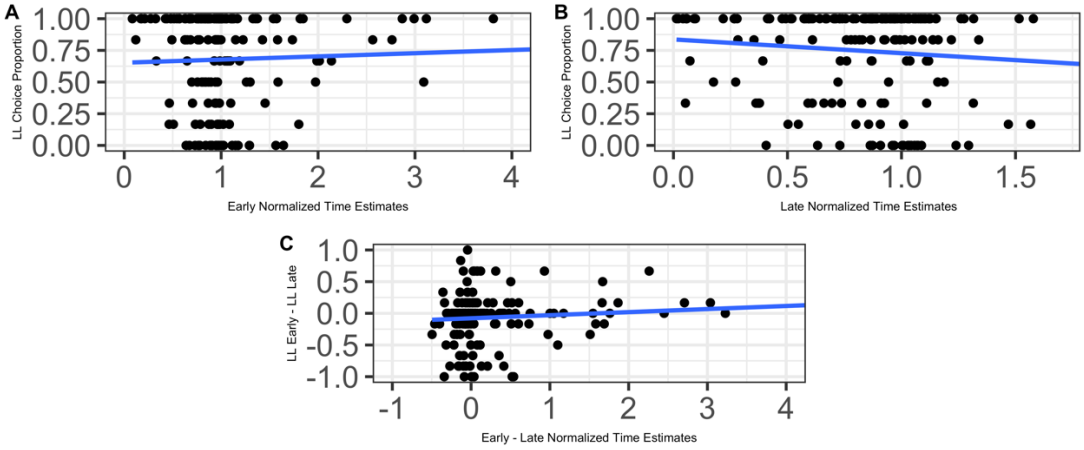

Figure 8: (A,B,C) The proportion of LL chosen against the mean normalised time estimate for each choice pair type, either early or late. Each dot shows the intersection of the proportion of LL choices for each participant and their mean normalised time estimate for each pair type. (C) The difference in LL chosen between early and late pairs against the difference between mean normalised time estimates for early and late pairs. Each dot shows the intersection of the difference in LL choice proportion and the difference in mean normalised times estimates for early and late pairs, for each participant. One outlier was excluded from Figure 8A.

Figure 8C plots the difference of LL chosen between the early and later pairs against the difference between the mean time estimates of early and late pairs. Again, there was no correlation between patience and time estimates $r(181) = -0.11$, $p = 0.14$. Given the presence of outliers we also conducted three exploratory Spearman Rank Correlations. For early pairs, there was no significant correlation between impatience and time estimates r_s (181) = -0.12, $p = 0.10$. There was also no correlation between time estimates and impatience for later pairs r_s (181) = -0.11, $p = 0.13$. However, some evidence for a correlation between the size of the common difference effect and time estimates r_s (181) = 0.18, $p = 0.018$ just above our pre-registered level of significance.

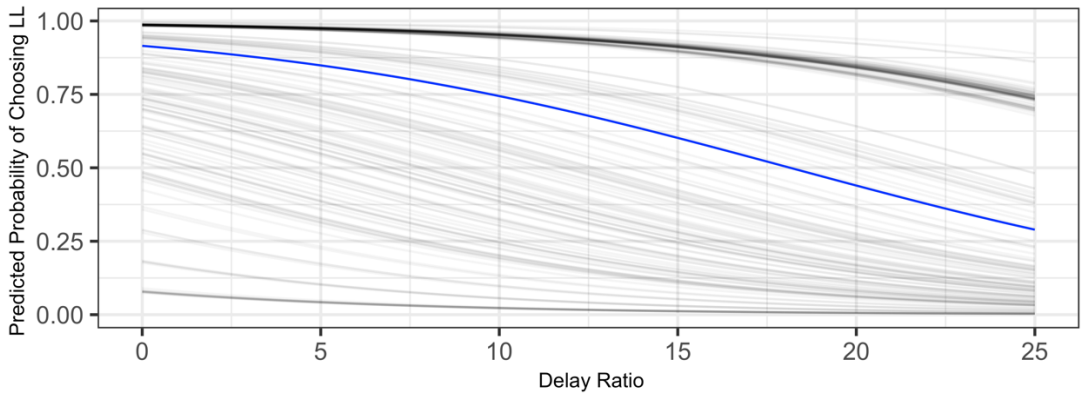

Figure 9: Predicted probability of choosing LL by delay ratio. The delay ratio is the mean time estimate of the late delays divided by the mean time estimate of the early delays. The blue line shows the predictions based on the fixed effect of delay ratio. The black lines show the predictions of the individual random effects. The darkness of the line indicates more participants.

One important finding in psychophysics is that individuals perceive ratios more saliently than differences (Stevens, 1975). Consequently, we conducted an exploratory analysis using the perceived ratios between the delays. A generalised linear mixed model with binomial family and logistic link function and by-participant random intercepts was used to estimate the effect of delay ratio on choosing LL. Here, the delay ratio is the mean time estimate of the late delays divided by the mean time estimate of the early delays. In Figure 9, as the ratio increases the probability of choosing LL decreases (Z = -5.03, p = <.001). One participant whose time estimate of a delay was >250 seconds was excluded from this analysis.

2.1.4. Discussion

In Experiment 1, we find a trend towards a common difference effect in intertemporal choice both in description and experience using delays on the order of seconds and consumable rewards. These findings are analogous to previous studies that have replicated classic behavioural effects in experience that are commonly found in description. Unlike decisions under risk, however, we did not find evidence for a description-experience gap in this aspect of decisions over time.

We posited that one potential mechanism for a description-experience gap in intertemporal choice could be a difference in time perception between description and experience. Specifically, we hypothesised that the size of the common difference in experience would be smaller as the difference between the delays in both the early pairs and late pairs would be smaller. We found no evidence for a differential effect on the way information was learned on the size of the common difference effect. This is surprising given that past efforts at replicating the common difference effect in experience typically find that the effect is smaller in experience than in description (Harman & Gonzalez, 2015).

However, we did find evidence for a difference in the accuracy of time estimates overall. Moreover, an exploratory analysis found a differential effect of condition on the accuracy of time estimates. We find some evidence for a correlation between time estimates and intertemporal choice, but not as strong as a relationship as predicted. In an exploratory analysis we did find that as the perceived ratio between the delays grew larger, the probability of choosing LL decreased. Moreover, we also find a central property of interval timing, time-scale invariance, in the data.

2.2. Experiment 2

2.2.1. Introduction

In Experiment 1, we found no evidence for a description-experience gap in intertemporal choice. This finding is in contrast to findings in decisions under risk where a well-known description-experience gap has been documented. We also did not find that the size of the common difference effect varied between description and experience. However, interestingly we found preliminary evidence for a common difference effect in experience. Lastly, we found little confirmatory evidence that individual differences in time perception correlate with intertemporal choices, but exploratory analyses did find that the perceived delay ratio did predict choice.

We also found preliminary evidence for differential time estimation between description and experience; however, the differences in time estimates did not

translate into different choices between description and experience. We hypothesised that the common difference effect would disappear in experience because of overestimating short delays and underestimating long delays. In the previous experiment, people underestimated large delays in experience, but did not overestimate short delays in experience. The lack of overestimation may have been why the common difference effect did not disappear in experience. In order to further elicit differential under and overestimating, in the present experiment we use a different set of delays as shown in Table 2. By using even shorter delays for the early delays set, we expect people will relatively overestimate the early delays even more in experience. Likewise, by using even larger delays in experience, we expect people to relatively underestimate the later delays. Using these new set of delays designed to further elicit over and underestimation, we will re-investigate the previous four key questions:

1) Do people differ in patience in D vs E?
2) Does a CDE appear (a) in experience and (b) with short delays?
3) Is there a CDE gap?
4) Does time perception relate to impatience?

Groups	**Delays (in S)**	
	Early	Later
Description	5 vs 15	85 vs 95
Experience	5 vs 15	85 vs 95

Table 2: Trial Types by Condition for Experiment 2

2.2.2. Method

2.2.2.1. Participants

Following our pre-registered sampling plan for Experiment 2, we collected a sample of 194 participants with 98 participants in experience and 96 participants in description (82 women; 83 men; 27 NA; Mage=22.35, SD=4.84). Only participants who passed both pre-registered exclusion criteria were counted. For this study there were two exclusion criteria, catch trials and time estimates. If participants failed

either of the exclusion criteria they were excluded. For the catch trials, the same catch trials were used as in Experiment 1 but with different delays. Participants who failed more than one catch trial were excluded. For the time estimates, participants whose mean time estimates were less than 5 seconds for 3 or more of the delays were also excluded.

Participants were paid a £3 show up fee and offered up to two pounds for performance. If a participant's time estimate was off by 20 per cent or less, then they were paid an additional 25 pence. Participants were recruited from the same pool as Experiment 1 but were excluded from the experiment if they participated in Experiment 1. Ethics approval was given by WBS NER committee 001/15-16. People arrived in the lab and were provided earphones and seated in front of individual computers. They were asked to place their phones and watches out of sight.

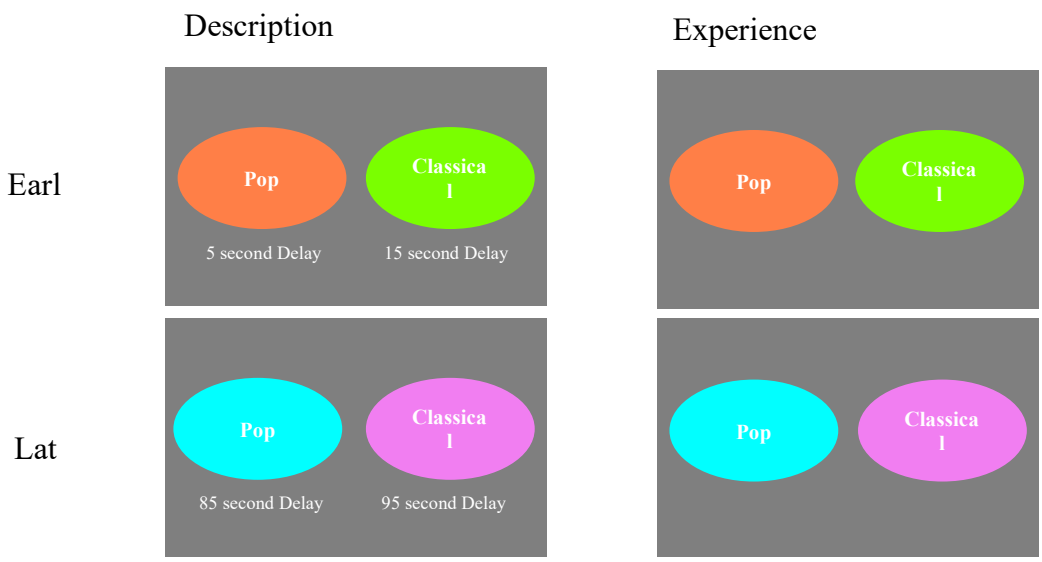

Figure 10: The columns show the between groups factor, description and experience, while the rows show the within groups factors, early and late. The rewards for each option are inscribed on each button, while the delay for each option is shown below in description. In experience they learn about the delays associated with each option (indicated by the colour) in a sampling phase prior to making their choices.

2.2.2.2. Design and Procedure

Experiment 2 used the same design and procedure as Experiment 1, except the length of the delays was changed. The between-subject factor was how knowledge of the delays was acquired with two levels: description or experience. The within-subject factor was delay length with two levels: early and late. There were four different delays: two early (five and 15 seconds) and two late (85 and 95 seconds) (Figure 10).

2.2.2.3. Data Analysis

Patience was calculated using the same method in Experiment 1. An identical pre-registered 2x2 repeated measures analysis of variance (ANOVA) with condition (description, experience) and delay length (early, late) was used to test whether there was differential patience between conditions and early and late trials. All degrees of freedom for one-way ANOVA tests were corrected using Greenhouse-Geisser correction**.** Our predefined level of significance was defined at $\alpha = .01$. All hypotheses, methods, and analyses were pre-registered at osf.io/fw9pg. Any exploratory analyses are marked as such in the analysis section.

2.2.3. Results

2.2.3.1. Description-Experience Gap and Common Difference Effect

Our first research question asked whether people would be more or less patient in description than in experience. People were not significantly more likely to select the LL option in experience than in description $F(1,192) = 0.17$, $p = .68$, $\eta_p^2 = 0.0009$ (Figure 11). Our second research question asked whether a common difference effect appears (a) in experience and (b) with short delays? In experience people were not significantly more likely to select the LL option with the later delays $t(97) = -0.19$, $p = .57$, $d = 0.02$ (Figure 12).

Our third research question asked whether there was a common difference experience gap between description and experience. People chose differently

between early and late delays between experience and description F(1,192) = 7.49, p = .007, η_p^2 = 0.04 (Figure 12). Furthermore, people were more likely to take the LL option in the late vs the early condition F(1,192) = 6.08, p = .01, η_p^2 = 0.03 (Figure 11B). In description, people were more likely to select the LL option with the later delays *t*(95) = -3.76, *p* = .0001, *d* = 0.38 (Figure 12).

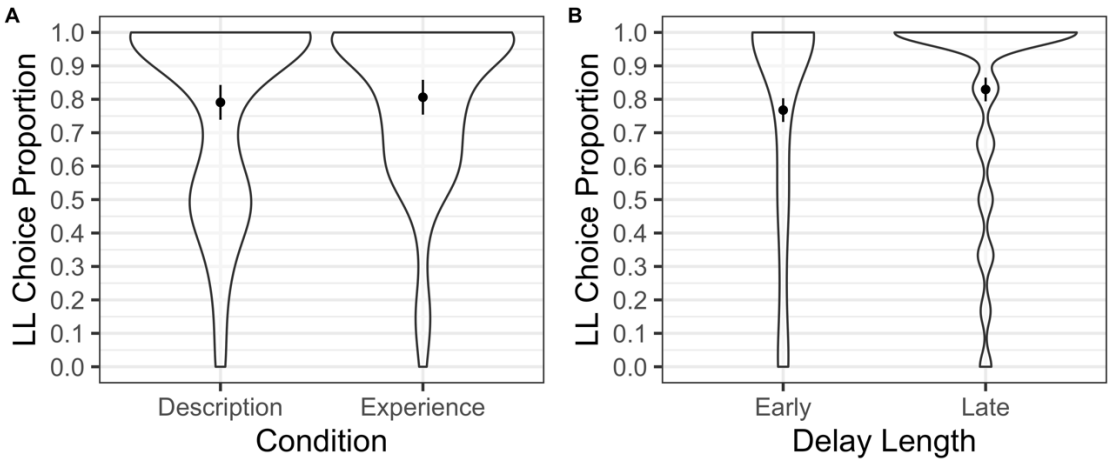

Figure 11: Proportion of LL choices on decisions trials between the description condition and experience condition. The dots show the mean choice proportion, while the bars indicate 95 per cent confidence intervals.

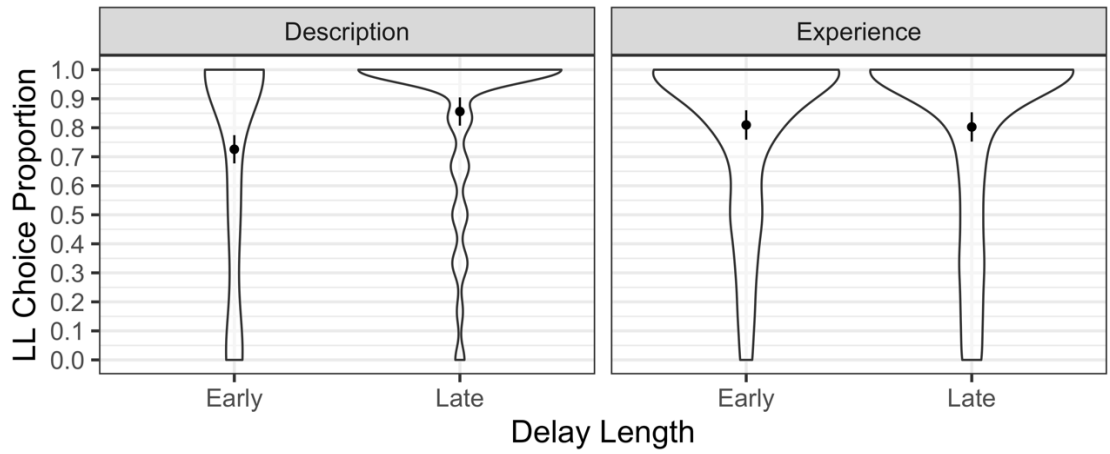

Figure 12: Proportion of LL choices on early and late trials by condition. The dots show the mean choice proportion while the bars indicate 95 per cent confidence intervals.

2.2.3.2. Time Perception

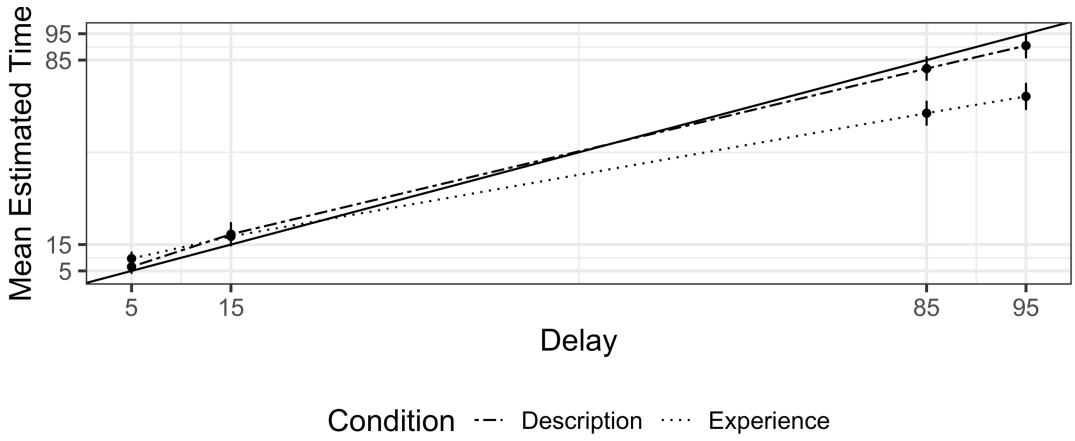

Figure 13:Mean time estimates for each delay by description and experience. The dots show the mean estimated time while the bars indicate 95 per cent confidence intervals within participants. The dotted line shows the experience condition while the dashed line shows the description condition. The solid line indicates perfect accuracy.

Figure 13 shows the mean time estimates for each delay. Again, a similar pattern of underestimation of late delays appear; however relative to Experiment 1 there appears to be more overestimation of early delays.

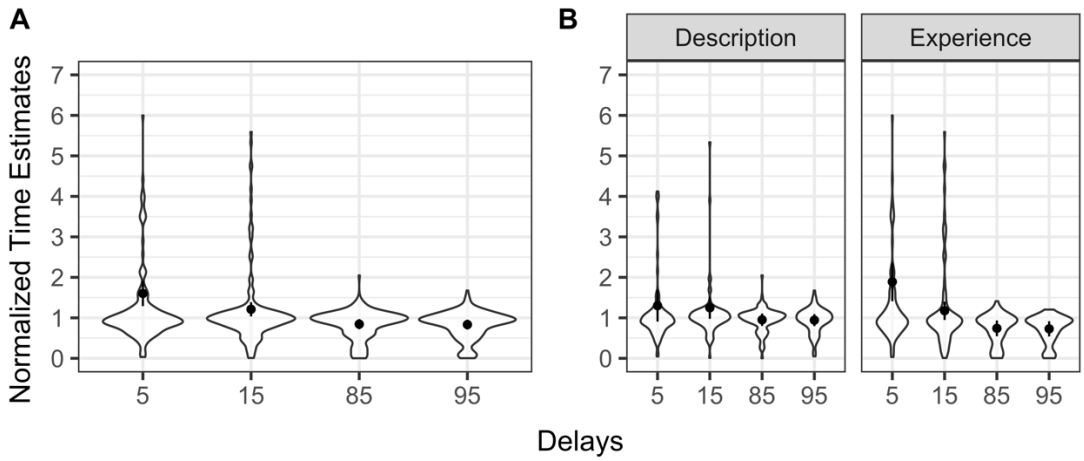

Figure 14: Mean normalised time estimates for each delay. The mean normalised time estimate is the mean of two time estimates divided by the true delay per participant. (A) The mean normalised time estimates collapsed across conditions. (B) The mean normalised time estimates for each delay by condition. A mean

normalised time estimate of one would indicate perfect accuracy. Nine observations were excluded from the plot for clarity.

Figure 14 plots the mean normalised time estimates for each delay collapsed across condition (Figure 14A) and separated by condition (Figure 14B). Similar to Experiment 1, overall the results indicate that there are differences in people's time estimates across delays and conditions. In the description condition participants are fairly accurate, while minimally overestimating the early delays and underestimating the later delays. In the experience condition, participants overestimated the early delay, while underestimating the later delays to a much greater extent. These visual differences were further confirmed by an exploratory one-way repeated measures ANOVA with mean normalised time estimates (mean of two estimates divided by the true delay) with delays (5, 15, 85, 95) and condition (description, experience). Overall, condition alone did not appear to influence estimates $F(1, 196) = 0.05$, $p = 0.82$, $\eta_p^2 = .06$; however, participants estimates for each delay differed $F(1.43, 279.63) = 12.88$, $p = <.0001$, $\eta_p^2 = .06$. Moreover, delays were differentially influenced by condition $F(1.43, 279.63) = 3.65$, $p = 0.04$, $\eta_p^2 = .02$, but not at our pre-defined level of significance.

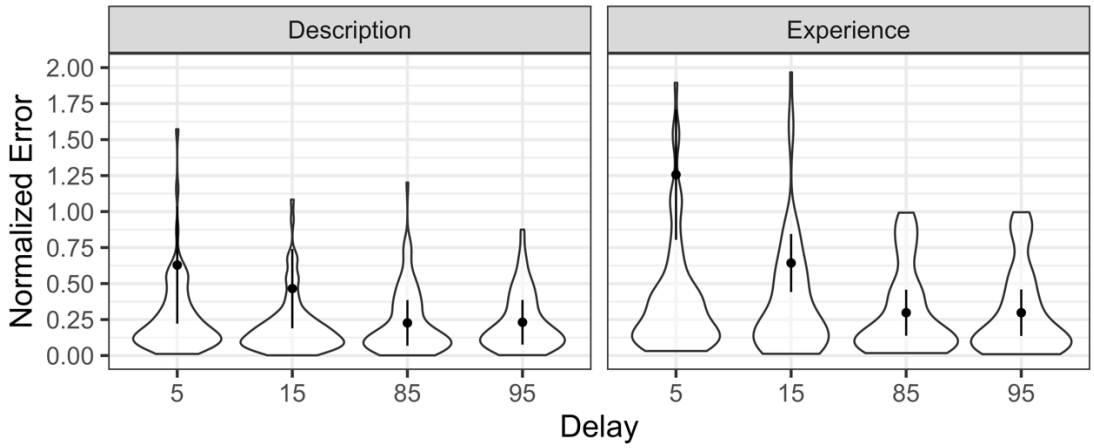

Figure 15: Normalised error for every delay for both description and experience. The normalised error is the absolute value of the delay subtracted by the time estimate divided by the delay. The dots show the mean normalised error, while the bars indicate 95 per cent confidence intervals.

Figure 15 shows the normalised error for each delay for the description and experience groups. Overall, there is a fair degree of scale invariance specifically for the later delays. However, the normalised error appears to be further bifurcated between short delays and longer delays than in Experiment 1. That is, the normalised errors appear to be higher for short delays in Experiment 1 than in Experiment 2. Similar to Experiment 1, the normalised error is higher for the five second delay than for the other delays. However, for the second shortest delay, 15 seconds, especially in experience, the error is much higher. One reason for this might be the delay differences between the two experiments. In Experiment 1, the delay difference is 20 seconds, while in the second experiment the delay difference is 10 seconds; thus, increasing the confusability of delays resulting in larger errors for the 15 second delay.

2.2.3.3. Choice and Time Perception

Our fourth question asked whether there was a relationship between time perception and choice. Figure 16A plots the frequency of LL chosen among early pairs against the mean normalised time estimate of early pairs. Similar to Experiment 1, there was no correlation between patience and time estimates $r(192) = 0.04$, $p = 0.51$ for early pairs Figure 16B plots the frequency of LL chosen among later pairs against the mean normalised time estimate of later pairs. There was a small correlation between patience and time estimates $r(192) = .19$, $p = 0.01$. Figure 16C plots the difference of LL chosen between the early and later pairs against the difference between the mean time times estimates of early pairs and late pairs. There was also a small correlation between patience and time estimates, $r(192) = 0.18$, $p = 0.01$. Given the presence of outliers in the dataset, we also conducted three exploratory Spearman Rank Correlations. For early pairs, there was no significant correlation between patience and time estimates r_s (192) = -0.01, $p = 0.86$. The size of the correlation between patience and time estimates for later pairs was similar to the Pearson correlation r_s (192) = 0.18, $p = 0.013$. In contrast to the Pearson correlations, there was no significant correlation between the size of the common difference effect and time estimates r_s (192) = 0.03, $p = 0.66$.

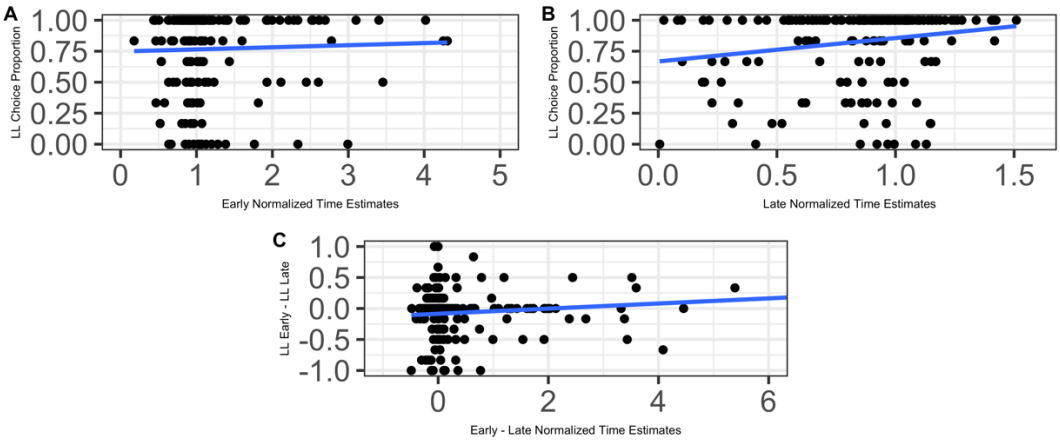

Figure 16: (A,B,C) The proportion of LL chosen against the mean normalised time estimate for each choice pair type, either early or late. Each dot shows the intersection of the proportion of LL choices for each participant and their mean normalised time estimate for each pair type. (C) The difference in LL chosen between early and late pairs against the difference between mean normalised time estimates for early and late pairs. Each dot shows the intersection of the difference in LL choice proportion and the difference in mean normalised times estimates for early and late pairs, for each participant. For clarity, four data points were excluded from Figures 16A and 16C.

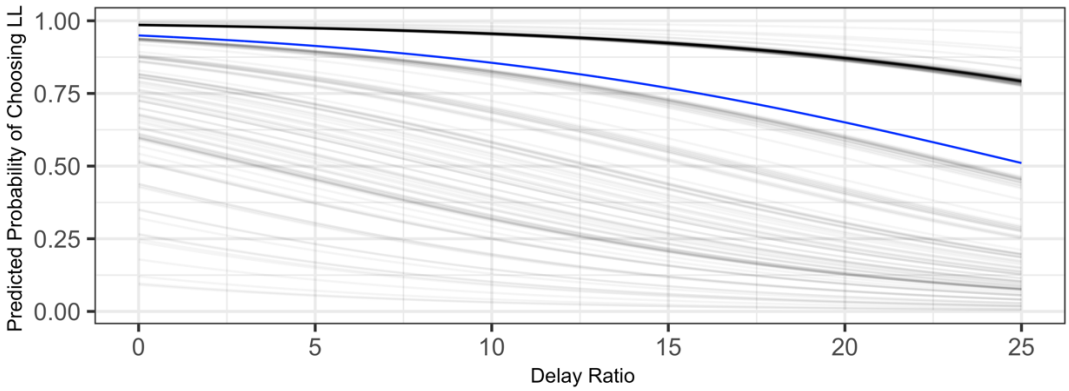

Figure 17: Predicted probability of choosing LL by delay ratio. The delay ratio is the mean time estimate of the late delay divided by the mean time estimate of the early delay. The blue line shows the predictions based on the fixed effects. The black lines show the predictions of the individual random effects. The darkness of the line indicates more participants.

The probability of choosing LL as a function of the delay ratio was estimated using an identical model in Experiment 1, a generalised linear mixed model with binomial family and logistic link function with by-participant random intercepts. Again, identical to Experiment 1 delay ratio is the mean time estimate of the late delays divided by the mean time estimate of the early delays. Figure 17 shows, similar to the results in Experiment 1 as the delay ratio increases the probability of choosing LL decreases (Z = -0.12, p = <.001).

2.2.4. Discussion

The results from Experiment 2 replicate the findings from Experiment 1, which shows a common difference effect using very short delays in description. In contrast with Experiment 1, the common difference effect disappeared in experience consistent with findings in risky choice that find a diminished common ratio effect in experience (Harman & Gonzalez, 2015). The results of Experiment 2 also indicate that there is no overall difference in patience between description and experience Taken together, the result from Experiment 2 show a common difference effect gap between description and experience, but no overall description-experience gap. Moreover, there was also mild evidence for a relationship between time perception and choice. When there was only underestimation of late delays in Experiment 1 the common difference effect appeared in experience, however when there was both under and overestimation in Experiment 2, the common difference effect disappeared entirely in experience.

In Experiment 2, the difference in delays between early and late trials was further increased. However, the effect size of the common difference effect remained the same. Unlike Experiment 1, overall the size of the common difference effect was driven by the description condition. One potential reason for this could be in Experiment 1 the difference between the delays was 20 seconds, while in Experiment 2 the difference between the delays was 10 seconds. In experience, as the difference in delays became less distinguishable, the rewards become more salient, meaning that participants chose their preferred reward regardless of the associated delay. This explanation would be in line with previous research in risk

that suggests that one reason for the common ratio effect is that when the alternatives become more similar, people simply choose their preferred reward (Leland, 1994; Rubinstein, 1988).

The role of time perception on choice across Experiments 1 and 2 is not entirely straightforward. The correlations in both Experiments 1 and 2 did not show strong evidence of a relationship between time perception and choice. However, in Experiment 1 there was an exploratory relationship between the perceived ratio between early and late delays (delay ratio) and a confirmatory relationship between the perceived ratio between early and late delays (delay ratio) in Experiment 2. In the time estimation task in Experiment 1, in the experience group, the five second delay is overestimated, while the rest of the delays are underestimated. In the time estimation task in Experiment 2, in the experience group, both the five and the 15 second delays are overestimated. There are potentially two explanations for why this might be. The first is that people overestimate short delays, meaning that 15 seconds could be considered short relative to 25 seconds. Alternatively, participants had a harder time discriminating between five vs 15 seconds than five vs 25 seconds. In the end, both potential reasons are inherently time perception based and appear to be the cause of the common difference effect disappearing in Experiment 2.

2.3. Experiment 1 and 2: General Discussion

In the domain of risk there is a well-known divergence between choices made from description and choices made from experience. This is known as the description-experience gap (Hertwig *et al.*, 2004; Hertwig & Erev, 2009; Wulff, Mergenthaler-Canseco & Hertwig, 2018). Only a handful of studies in the domain of time have sought to investigate whether a description-experience gap occurs in intertemporal choice with mixed results (Dai *et al.*, 2019; Lukinova *et al.*, 2019). In the present study, we ran two experiments investigating whether such a gap exists in intertemporal choice. Overall, we find conflicting evidence for a description-experience gap in intertemporal choice. In Experiment 1 we found both a common difference effect in description and experience, which suggests that there is no common difference effect gap in intertemporal choice. However, in Experiment 2,

and consistent with our predictions, the gap disappeared in Experiment 2 — suggesting a common difference effect gap does exist in intertemporal choice.

There could be a number of potential reasons for not consistently finding a description-experience gap in intertemporal choice. Most studies that find the description-experience gap in decisions under risk use choices with a rare event (but see Ludvig & Spetch, 2011). Similarly, Dai *et al*. (2019) find a description-experience gap in intertemporal choice, specifically that people underweight the probability of encountering the rare delay in experience and overweight the probability of encountering the rare delay in description. In Experiments 1 and 2, there is no rare event nor are any of the delays probabilistic. In any intertemporal choice there is of course some uncertainty as to whether the delay will arrive, potentially adding probabilistic delays by definition increases uncertainty. Moreover, adding probabilistic delays may also make the uncertainty inherent in intertemporal choice more salient similar to how the hidden zero effect (Magen, Dweck & Gross, 2008; Read, Olivola & Hardisty, 2016) makes trade-offs more explicit.

In Experiment 1, we find a trend towards a common difference effect in both description and experience. Although, in Experiment 2 we only find a common difference effect in description. One reason for the selective appearance of the common difference effect may be due to difficulty in discriminating between delays, which introduces uncertainty. This in turn may cause people to focus on their reward preferences rather than their time preferences.

To our knowledge this is the first study to find the common difference effect both in experience and on the order of seconds rather than days, months, or years. Moreover, we also find a common difference effect using real consumable rewards. Overall, finding the common difference on the order of seconds is surprising given the length of the delays previously used. One potential reason for the appearance of the common difference effect on the order of seconds in description is that people do not pay attention to the units but rather only the numbers (Furlong & Opfer, 2009). This is similar to the way attribute-based models of intertemporal choice suggest people make decisions over time. However, this explanation does not entirely suffice

given the appearance of the common difference effect in experience where no units or numbers are presented.

Chapter 3. Uncertainty and the Common Difference Effect

3.1. Introduction

The results of Experiment 1 show a common difference effect in both description and experience. However, in Experiment 2 we only find a common difference in description. The key difference between these two experiments was that the difference between delays for the smaller-sooner (SS) reward and the larger-later (LL) reward was 20 seconds in Experiment 1 and only 10 seconds in Experiment 2. We hypothesised that the reason for not finding the common difference effect in experience in Experiment 2 was that distinguishing between the delays in Experiment 2 was more difficult than in Experiment 1, and this in turn introduced more uncertainty. Consequently, people simply focused more on the reward than the delay because of the uncertainty, thus diminishing the size of the common difference effect. As a result, in both experience conditions they chose the larger reward 75-80 per cent of the time, similar to the proportion chosen with the later delay in description.

These findings suggest that by making the delays harder to discriminate between, people will become more patient, because instead of focusing on time preferences, they focus on reward preferences. Indeed, this finding holds across experiments. In Experiment 1, where the difference between the delays was 20 seconds, people chose LL around 70 per cent of the time. In Experiment 2, where the difference between delays was 10 seconds, people chose LL around 80 per cent of the time.

Similar to previous experiments investigating time ambiguity (Ikink *et al.*, 2019), in this Chapter, we attempt to resolve this puzzling pattern of results by isolating differences in time perception between description and experience. In Experiments 1 and 2, people appear to have been uncertain about the delays as shown by the normalised errors in each experiment which were around 40 per cent. Here, in Chapter 3, we attempt to specifically isolate whether time perception is the underlying determinant of the uncertainty around the delays (as shown by the time

estimation data) in Experiments 1 and 2, particularly when the delays that are being compared are similar in duration.

One way to increase uncertainty for described delays is to give a range for how long a delay will be rather than a fixed duration. This additional uncertainty in description resembles experience more as the exact delays are unknown when learned from experience. For example, if the normalised error for a delay of 25 seconds is around 40 per cent then participants' mental representation of the delay in experience may look like 15s-25s. Thus, by making the mental representation of delays in experience explicit in description, we isolate if time perception is a reason for differences between description and experience in intertemporal choice.

For the first research question we attempted to replicate the common difference effect again. (Common Difference Effect Replication). We constructed an identical description condition to Experiments 1 and 2; however, we utilised more choice pairs as shown on page 48. Based on past research (Keren & Roelofsma, 1995; Kirby & Herrnstein, 1995; Green & Myerson, 2004; Read *et al*., 2005; Scholten & Read, 2006; Holt *et al*., 2008) and the results from Experiments 1 and 2, we will attempt to replicate our findings from Experiments 1 and 2 by reproducing the common difference effect with very short delays.

The second research question (Uncertainty-Reward/Ambiguity-Pessimism) consists of two opposing predictions. By introducing more uncertainty into the range condition people could react in one of two ways. Firstly, people could focus more on the rewards rather than the delays and simply choose whichever option is associated with their preferred reward regardless of the potential delay (Uncertainty-Reward). Conversely, when making decisions under uncertainty people are often ambiguity averse (Trautmann & van de Kuilen, 2015), and choose as though they are pessimistic about the possible outcomes. Thus, in the range condition people will be more pessimistic and focus on the upper end of the delays. As a result of being pessimistic about the delays, people will believe that their preferred reward is farther into the future and consequently they will select the sooner rewards more often across all conditions with ranges (Ambiguity-Pessimism).

For our third research question (Delay Difference), we aim to understand how the differences in delays moderate uncertainty. Participants showed a common difference effect when the delays were 20 seconds apart (Experiment 1), but not when they were 10 seconds apart (Experiment 2). In both Experiments 1 and 2, they showed a common difference effect in description. If additional uncertainty is the key variable that distinguishes the conditions, then people will show a CDE with the wide delays (20 seconds apart), but not with the narrow delays (10 seconds apart). With the narrow delays, they will only pay attention to the outcome and will tend to select the Larger Later outcome more often for both early and late delays. This hypothesis predicts that the CDE will be similar for three conditions (Point-Wide, Point-Narrow, and Range-Wide), but will be smaller or even non-existent for the final condition (Range-Narrow). Accordingly, our research questions are as follows:

(1) Common Difference Effect Replication: Replicate our findings from Experiments 1 and 2 by reproducing the common difference effect with described point estimates with very short delays.

(2A) Uncertainty-Reward: Main effect of condition, with more patience in the range than the point estimate condition for all delays.

(2B) Ambiguity-Pessimism: Opposite prediction to 2A, less patience in range condition for all delays.

(3) Delay Difference: The additional uncertainty may only influence delays that are sufficiently close to each other.

3.2. Methods

3.2.1. Participants

In accordance with the pre-registration plan we aimed to collect 192 participants with 96 participants in each condition. Each experimental session was

run with up to 30 participants at a time. Ultimately, we collected a sample of 140 participants in the point group and 153 participants in the range group. The only participants included in the analysis were those who passed the pre-specified exclusion criteria, which is specified in the section Phase 2. After applying the exclusion criteria, this left 99 participants in the point condition and the 97 participants in the range description (93 women; 87 men; 16 NA; Mage = 20.43, SD = 2.89). Participants were paid a £3 show up fee and were able to earn up to an additional £1.60 based on their performance. Participants were recruited from the same pool as the first two experiments. The sample size is based on a power analysis for a two-tailed between-participant t-test with a medium effect size (d = 0.5), alpha = .01 to account for multiple comparisons, and power = 0.8. For the one-tailed within-participant tests (see below), this sample size provides 80 per cent power to detect an effect size of 0.23 with an alpha of .01. Power and sample sizes were calculated using G*Power 3.1. Participants who participated in Experiments 1 and 2 were not eligible to participate in the current study.

3.2.2. Design and Procedure

The study was a 2x2x2 mixed design. The between-subject factor was how the delays were presented with two levels: point estimates or ranges. The first within-subject factor was delay length with two levels: early and late. The second within-subject factor was delay difference with two levels: wide and narrow. There were eight unique delays, which were presented as point estimates or as ranges (Table 3).

Groups	**Delays**			
	Early-Narrow	Early-Wide	Late-Narrow	Late-Wide
Points	10 vs 20	5 vs 25	90 vs 100	85 vs 105
Ranges	6-14 vs 12-28	3-8 vs 15-35	54-126 vs 60-140	51-119 vs 63-147

Table 3: Trial Type by Condition for Uncertainty and the Common Difference Effect

Phase 1: Reward Selection. Participants viewed two playlists: one composed of classical songs and the other composed of pop songs. They then selected which playlist they preferred to randomly hear a song from.

Phase 2: Decisions. Next, participants made several choices between rewards at different delays. Options were displayed as coloured buttons with the potential music reward inscribed on them. The potential delay for each option was displayed underneath the button. Participants made a total of 28 choices. Of these, 24 were decision trials, which pit a smaller, sooner reward against a larger, later one. The trials were either early trials — where the sooner reward comes in either five seconds or 10 seconds and the later reward in 25 or 20 seconds — or late trials — where the sooner reward comes in 85 or 90 seconds and the later reward in 105 or 100 seconds (for each pair listed the time for the narrow delay difference comes first). Participants were shown every pair six times, with 24 decision trials in total. For each participant, their preferred playlist was paired with the larger delay and their non-preferred playlist was paired with the smaller delay. The side that the smaller delay and the larger delay appeared on was randomised. In addition to the decision trials, there were four catch trials. There was one catch trial for each choice pair, where the rewards were identical, but the delays were different. If participants chose the option with the longer delay more than once, they were excluded from the analysis.

Phase 3: Time Estimation. In this phase, participants made two separate duration estimates of the delays they chose between in the decision phase twice. In order to record each time estimate, participants were randomly shown each delay that appeared in the decision phase twice. A coloured button, representing the delays, appeared on the screen one at a time. In the range condition, underneath each button there was a description of the delay range. In the point condition, a description of the delay appeared underneath each button. For both conditions, after being shown the coloured button, participants pressed the space button to start the timer and then space again to stop the timer. In the range condition, for each delay, if a participant's estimate was within the range, they were rewarded 10 pence. Participants could earn up to an additional £1.60 from the 16 time estimation trials

with the possibility of earning up to £4.60 (£3 show up + £1.60 performance). In the point condition, for each delay, if a participant's estimate was within 40 per cent (which is the same width as the range) they were similarly rewarded.

Phase 4: Musical Reward. One of the participant's choices from Phase 2 was randomly selected. Participants then waited for the corresponding delay to earn their musical reward.

3.2.3. Data Analysis

For each condition patience was computed as the proportion of Larger Later (LL) choices made on the decision trials. LL is defined as the option with the larger reward and larger delay. Furthermore, patience was calculated separately for early and late and short and wide options to quantify the size of the common difference effect and whether it holds across delay differences. A pre-registered 2x2x2 repeated measures analysis of variance (ANOVA) with condition (range, point), delay length (early, late), and delay differences (narrow, wide) tested whether there was differential patience between conditions and early and late trials and short and narrow trials. All degrees of freedom for one-way ANOVA tests were corrected using Greenhouse-Geisser correction**.** Our predefined level of significance was defined at $\alpha = .01$. The hypotheses, methods, and analyses were pre-registered https://osf.io/zjn3t/. Exploratory analyses are marked as such in the analysis section.

3.3. Results

3.3.1. Uncertainty and the Common Difference Effect

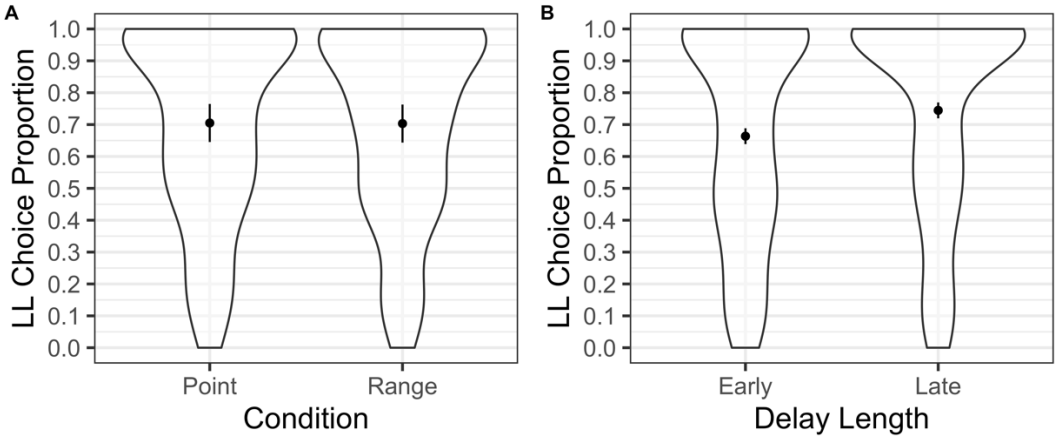

Figure 18: Proportion of LL choices on decisions trials between the description condition and experience condition. The dots show the mean choice proportion while the bars indicate 95 per cent confidence intervals.

Our first research question asked whether we again can replicate the common difference effect. Similar to Experiments 1 and 2, people were more likely to take the LL Option in the late condition than in the early condition $F(1,194) = 20.60$, $p < .0001$, $\eta_p^2 = 0.10$ (Figure 18B). Our second research question involves two competing predictions. Firstly, the Uncertainty-Reward predictions, which predicts that people will be more patient in the range condition. Secondly, the Ambiguity-Pessimism, which predicts that people will be less patient in the range condition. People in the Range condition were about equally likely (0.18±11.95%) to select the LL Option as in the Point Condition $F(1,194) = 0.06$, $p = .97$, $\eta_p^2 = <.0001$ (Figure 18A). In the points condition people were more likely to choose the LL option with the later delays $t(197) = -3.15$, $p = .001$, $d = 0.22$. Likewise, in the range condition people were more likely to choose the LL option with the later delays $t(193) = -4.05$, $p < .001$, $d = 0.29$ (Figure 19).

Our third research question investigated whether additional uncertainty only influences delays that are sufficiently close to each other. For delay differences, participants in the narrow condition were more likely to select the LL option in the

narrow vs wide condition $F(1,194) = 56.6$, $p < .0001$, $\eta_p^2 = 0.23$. The effect of delay difference by condition trended towards significance $F(1,194) = 3.09$, $p = .08$ $\eta_p^2 = .02$. Moreover delay difference varied by delay length $F(1,194) = 7.78$, $p < .006$, $\eta_p^2 = .04$. Delay difference did not vary across delay length and condition $F(1,194) = 0.46$, $p < .50$, $\eta_p^2 = 0.002$ (Figure 20) .

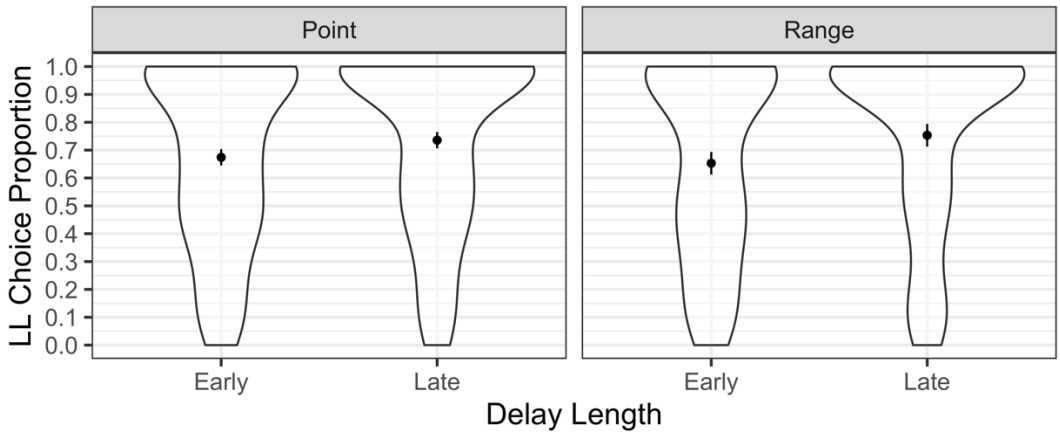

Figure 19: Proportion of LL choices on early and late trials by condition. The dots show the mean choice proportion while the bars indicate 95 per cent confidence intervals.

People chose the LL option more often in the point narrow condition $t(98) = -1.57$, $p = .06$, $d = 0.16$ as well as in the point wide condition $t(98) = -2.75$, $p = .004$, $d = .28$. Similarly, in the range narrow condition people chose the LL option more often in the range narrow condition $t(96) = -1.90$, $p = .03$, $d = 0.19$ as well as in the range wide condition $t(96) = -3.65$, $p < .001$, $d = 0.37$.

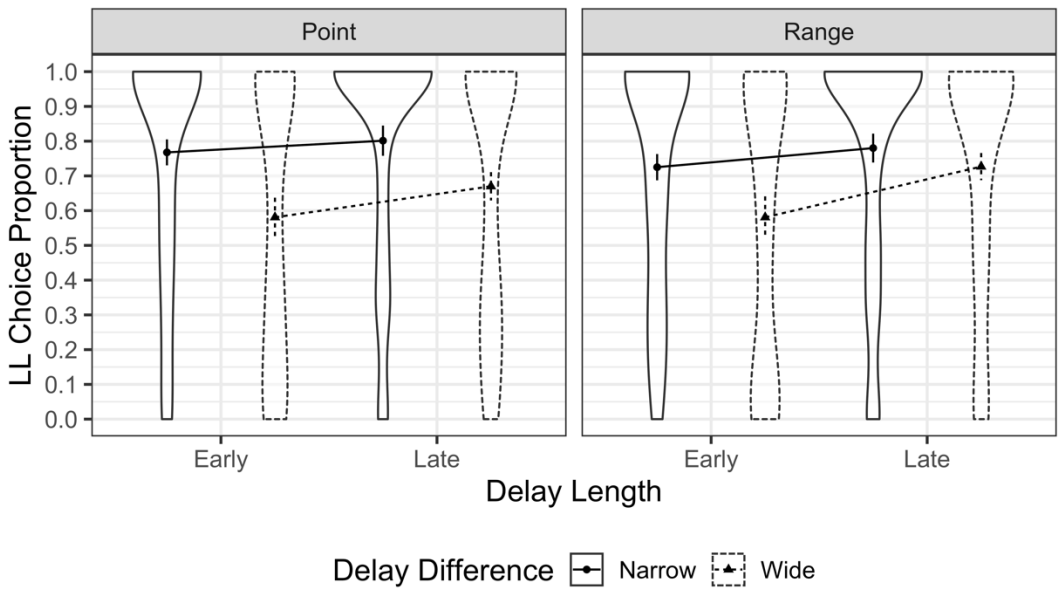

Figure 20: Proportion of LL choices on early and late trials by condition and delay difference. The dots show the mean choice proportion, while the bars indicate 95 per cent confidence intervals.

3.3.2. Time Perception

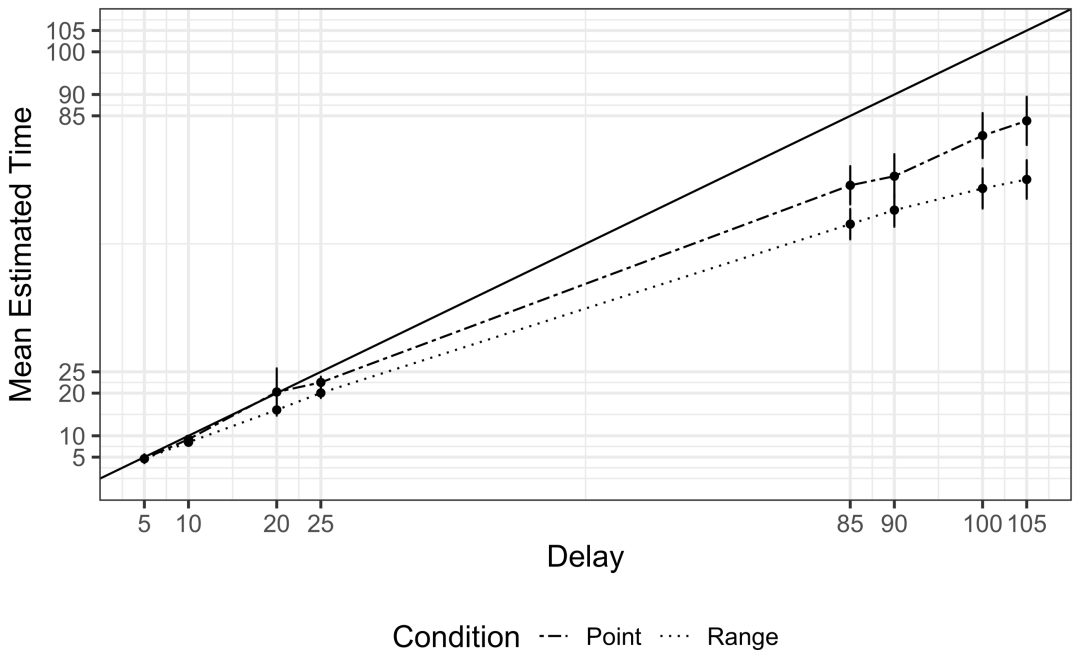

Figure 21: Mean times for each delay by points and range condition. The dots show the mean estimated time while the bars indicate 95 per cent confidence intervals.

The dotted line shows the range condition while the dashed line shows the point condition.

Figure 21 shows the mean time estimates by each delay between the Point and Range conditions. For the early delays peoples' time estimates are very close to the true time. However, similar to Experiments 1 and 2, the late delays are systematically underestimated.

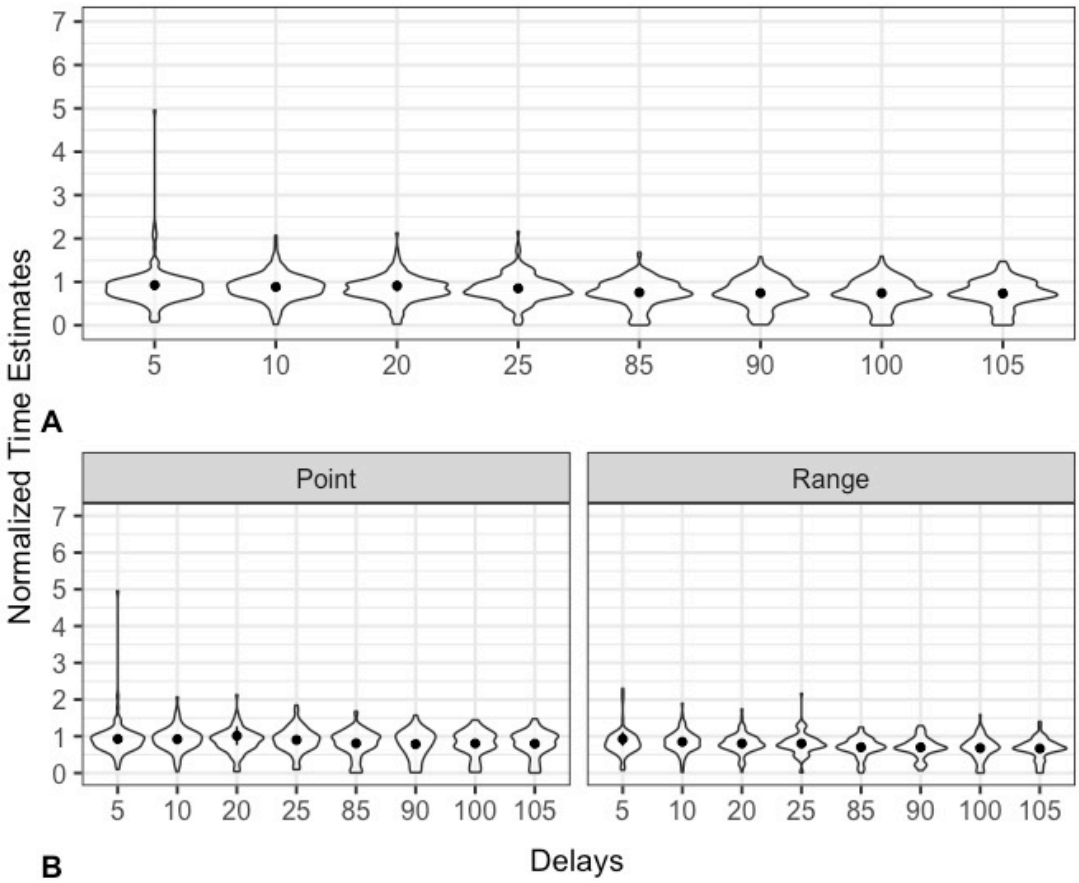

Figure 22: Mean normalised time estimates for each delay. The mean normalised time estimate is the mean of two-time estimates divided by the true delay per participant. A mean normalised time estimate of one would indicate perfect accuracy. (A) The mean normalised time estimates collapsed across the points and ranges groups. (B) The mean normalised time estimates for every delay by each group. Three outliers were removed from the plot for clarity.

Figure 22 plots the mean normalised times overall (Figure 22A) and by group (Figure 22B). Figure 22 suggests that there are differences in people's time estimates of the different delays. However, the differences between groups appear

negligible. These graphical comparisons were confirmed using an exploratory one-way repeated measures ANOVA with mean normalised time estimates (mean of two estimates divided by the true delay) with delays (5, 10, 20, 25, 85, 90, 100, 105) and condition (Point, Ranges). There was a trend towards a difference in people's time estimates between conditions $F(1,194) = 5.77$, $p = 0.02$, $\eta_p^2 = .04$. Moreover, there was a difference in people's time estimates between delays $F(2.27, 440.33) = 6.18$, $p = <.001$, $\eta_p^2 = .03$. Furthermore, people's time estimates of delays did not differ across conditions $F(2.27, 440.33) = 0.79$, $p = 0.47$, $\eta_p^2 = .004$.

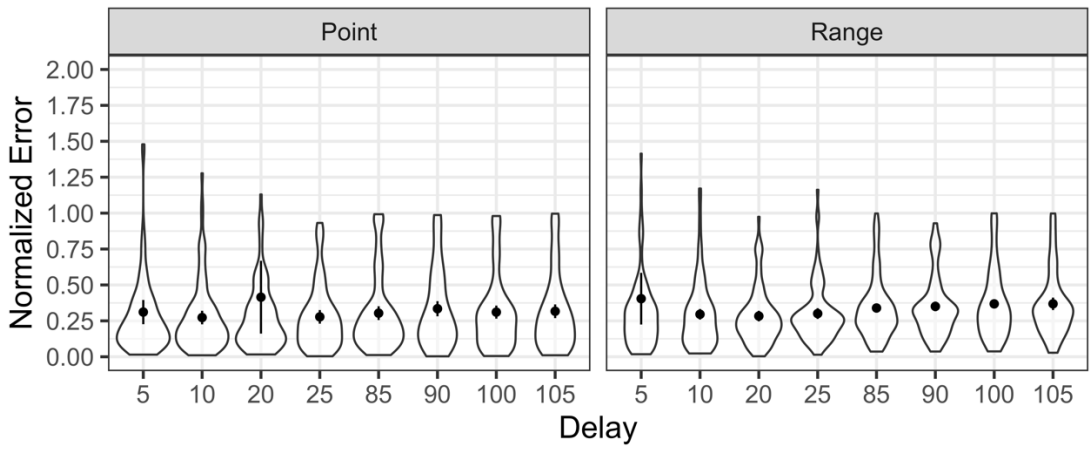

Figure 23: Normalised error for every delay by points and ranges. The normalised error is the absolute value of the delay subtracted by the time estimate divided by the delay. The dots show the mean normalised error, while the bars indicate 95 per cent confidence intervals.

Figure 23 depicts the normalised error for the points and ranges conditions. Similar to the experiments in Chapter 2, the mean for each delay sits between .25 and .50. Again this suggests that people's time estimates are scale invariant (Gallistel and Gibbon, 2000).

3.4. Discussion

Based on the results of Experiments 1 and 2, we hypothesised that uncertainty would modulate the size of the common difference effect. In the points group, we presented delays as points, which was meant to recreate the description

conditions in Experiments 1 and 2. In the range group, we presented delays as ranges, which was meant to parallel the uncertainty in the experience condition in Experiments 1 and 2. We also hypothesised that another source of uncertainty could be from delay differences. Accordingly, we also included choice pairs similar to Experiment 1 where the delay difference was 20 seconds and choice pairs similar to Experiment 2 where the delays difference was 10s.

Consistent with Experiments 1 and 2 we again find a common difference effect using delays with seconds rather than months days or year. We found no overall difference in patience when the delays are described as exact points in time or described as ranges. However, increased uncertainty, manifested in narrow delay differences, appeared to mildly reduce the size of the common difference effect as predicted according to our Delay Difference prediction. Delay differences between choice pairs did matter; participants were more patient for narrow delays than for wide delays. However, the effect of delay differences, combined with the uncertainty of the delays did not significantly change choice behaviour across groups.

Our results in the context of previous intertemporal choice studies have a number of implications for future studies of the description-experience intertemporal choice. Ikink *et al.* (2019) find that when uncertainty is added to either LL or SS, participants are less likely to choose the ambiguous vs the certain alternative, exhibiting typical ambiguity aversion. Dai *et al*. (2019) find a divergence in choice when participants choose between uncertain and certain alternatives. In the present study, participants either only see uncertainty paired with both Smaller Sooner and Larger Later options in the ranges condition or only certain Smaller Sooner and Larger Later options in points condition in contrast to studies in probability space (Dai *et al.*, 2019) and time space (Ikink *et al.*, 2019). Accordingly, one possible reason we do not observe a consistent gap between groups in the experiment is that participants do not explicitly trade-off between certain and uncertain alternatives. Studies that suggest there is no description-experience gap in intertemporal choice do not have participants explicitly choose between certain and uncertain alternatives (Lukinova *et al.*, 2019).

Other studies examining the intersection of choice involving risk and time (risky inter-temporal choices) find that people's risky and intertemporal choices are captured best by a unitary model that incorporates both risky and intertemporal choices (Luckman, Donkin & Newell, 2018). This suggests that both risk and intertemporal choices are simply special cases of an overarching category, risky intertemporal choices. According to this framework, intertemporal choices studies that do find the gap may simply be closer to risk on the risk intertemporal choice spectrum.

3.5. Interim Summary

In Experiment 1, we found a common difference effect in description and experience using very short delays (seconds). In the second experiment we sought to replicate the results of Experiment 1, using choice pairs that only had differences of 10 seconds rather than 20 seconds. Again, people exhibited the common difference effect in description, although not in experience. One explanation for the selective appearance of the common difference effect could be due to uncertainty around the length of the delays in experience. We tested this hypothesis in Experiment 3 and found inconclusive evidence for the role uncertainty plays. We did find that the difference between delays is influential for choice; however, the effect of delay difference only varied mildly depending on whether the descriptions of the delays were presented as points or ranges.

Our findings have a number of implications for intertemporal choice. Firstly, unlike the expansive literature in risk documenting a description-experience gap, we find inconsistent evidence for a description-experience gap in intertemporal choice in the form of the common difference effect. There are a number of potential reasons why this might be. Studies examining the description-experience gap almost always focus on underweighting rare events in description and overweighting rare events in description (but see Ludvig, Madan, McMillan, Xu, & Spetch, 2018; Ludvig & Spetch, 2011). In all three of our experiments, there is no rare event. Thus, our findings may suggest that a consistent description-experience gap may not occur in intertemporal choice without rare events.

A potential design concern with many intertemporal choice studies is the effect of anticipatory utility (Loewenstein, 1987). That is, people might derive utility by delaying the consumption of their preferred reward. In effect, this would mean that people would not be making the expected trade-off between delay and reward. In Experiment 3, we control for this possibility by including catch trials where the delays are different, but the rewards are the same. If participants chose the alternative with the larger delay on more than two out of the four catch trials, they were excluded from the experiment thus ensuring that only participants who preferred larger rewards sooner were included in the analysis. While the same catch trials were not used in Experiments 1 and 2, similar results were observed across all three experiments, which suggests that anticipatory utility did not have an outsize effect on the results.

Outside of using experience-based choices, one way in which our study builds on prior work on intertemporal choice is the length of delays. Many studies use choices on the order of days, months, and years (Frederick *et al.*, 2002) rather than seconds. Despite the short delays, participants still choose the Smaller Sooner alternative in our experiment, which is consistent with other recent experiments that have used delays on the order of hours (Augenblick, 2017) and even seconds (Ashby & Gonzalez, 2017).

Taken together, our results show that intertemporal choice phenomenon in description can also be found in experience. Moreover, our results demonstrate that participants exhibit significant discounting behaviour even over very short delays. Interestingly, we did not find consistent evidence for a description-experience gap in intertemporal choice. Future research on the description-experience gap on decisions over time could utilise different sampling paradigms as we only examined one potential variant. Furthermore, in the current design, participants only experienced the delays before making their choice. Examining, whether experiencing both the delays and rewards will produce different behaviour could also be a fruitful endeavour.

Chapter 4. Tests of Social Comparison-Based Feedback

4.1. Introduction

How to provide the best feedback on academic performance is an unsolved question. One of the simplest and most widely used forms of feedback given is a simple numerical score. This numerical score is often accompanied by additional information such as how the score compares to the mean or where the score lies in the distribution of scores (the rank). Rank-based feedback has been used in experiments to encourage information seeking of alcohol related health information (Taylor, Vlaev, Maltby, Brown, & Wood, 2015), increase the willingness to pay for health foods (Aldrovandi, Brown, & Wood, 2015), and increase teeth brushing (Maltby *et al.,* 2016). This non-educational stream of research has generally found positive effects of rank. In contrast, in the domain of education the effects of social comparison such as providing rank-based feedback have been more mixed.

In the present study we draw directly on the frames used in Taylor, Vlaev, Maltby, Brown, and Wood (2015). In their study they sought to reduce alcohol consumption through providing either rank, mean, or absolute feedback. In the absolute condition they provided excessive alcohol drinkers with official health guidelines for drinking. In the mean condition, they provided excessive drinkers with the mean amount of drinks of participants in the sample with the same gender. In the rank condition, participants were told how their self-reported units per week consumption ranked in the sample. Overall, participants in each condition reported using less alcohol on the follow up questionnaire than on the baseline questionnaire. Furthermore, there was no difference in self-reported alcohol consumption between conditions. Importantly, participants in the rank condition were more likely to request information on alcohol counselling services compared to other conditions.

Similar to Taylor, Vlaev, Maltby, Brown, and Wood (2015), we experimentally manipulated whether students in an inverted classroom received either a rank, mean, or absolute feedback reminder. The goal of the current study

was to provide a direct test of the efficacy of rank vs mean-based feedback on academic performance. Consistent with our predictions, rank was the most effective feedback reminder followed by mean and absolute feedback reminders. However, these results were not statistically significant. A secondary goal of our study was to also increase information seeking about students' own grades. In addition to one of the frames, students were provided a link with further information about their grade. Conversely, students who received the absolute treatment were more likely to view additional information about their scores followed by the mean and rank treatment.

In the sections below, various mechanisms behind social comparisons are discussed. Moreover, traditional measures of effort and academic success are reviewed. Additionally, how social comparisons have been used in previous experiments — both natural and experimental — are examined. Lastly, the methodology and results of the experiment are presented, and the results are discussed as well as limitations.

4.2. Lab and Field Intertemporal Choices

In the previous chapters, a series of lab-based experiments was described that involved two key dimensions; intertemporal choice and experience. A reason for incorporating experience into intertemporal choice is to increase ecological validity of the task itself as often choices in everyday life involve relying on one's experience. Another way of increasing ecological validity is by examining peoples' behaviour in the field.

A common scenario where people need to decide how to allocate their time based on feedback is when studying at university. Imagine a student is given multiple assignments across several different classes. After each assignment the student is given feedback and needs to decide how much time they need to allocate studying for each class based on the feedback they receive. In this chapter, a randomised control trial is introduced where students need to learn to calibrate their effort based on the feedback they receive. This RCT can be conceived as a real-world test of how people make trade-offs based on their past experiences.

4.3. Mechanisms for Change: Informational vs Non-Informational Influences on Effort

There are several mechanistic explanations as to why rank or mean information might be impactful in an educational context. One mechanism might be how objectively useful is the information itself, and secondly, how does that information affect the student. For the objective usefulness of the information, the educational context is important. In the US educational system, often times students are graded on a curve. A student could score 40 per cent on an assessment; however, this score could still be curved to 100 per cent depending on the performance of the other students. Assuming that there are a large number of students and normally distributed scores, rank information is much more valuable information than the mean in the US context because a student's score is determined by their peers' performance. In the UK educational system, rank information is not as informative because students are not graded on a curve. Therefore, rank is not nearly as important because one's grade isn't determined by other students' performance.

Students of different abilities might react differently to information communicated by relative vs absolute grading. Becker and Rosen (1992) propose a model that shows when high ability students are evaluated using absolute standards their efforts increase. Conversely, when high ability students are evaluated using relative standards, they lower their effort. Other theoretical results show that under all circumstances students should exert more effort under an absolute grading scheme (Dubey & Geanakoplos, 2010). Using empirical evidence from Chile in which the grading scheme was switched from absolute to relative grading, Paredes (2016) shows that switching grading schemes affected student effort as measured by their grades. When the grading scheme was absolute, high ability students exerted more effort, while low ability students exerted less effort. When the grading scheme switched from absolute to relative, low ability students exerted more effort, while high ability students exerted less effort.

Another mechanism might be how rank affects motivation or the self-concept of students. If a student is ranked first but is only in a class of two people, this is not

as informative as when there are 500 other students, yet coming in last place could affect a student's motivation or self-concept (Gill *et al.*, 2018). Some studies have shown that these non-informational effects can have an outsize influence. In the UK system where only absolute performance is measured, Murphy and Weinhardt (2018), using a sample of two million students from England, found that rank had a significant effect on later academic achievement. Moreover, boys gain four times more in later tests scores as a result of being higher ranked relative to girls. They attribute the gains in performance due to a high rank improving self-confidence and belief in one's own ability. Similarly, the non-informational influence of rank has also been observed in the US. Elsner and Isphording (2016) compared multiple cohorts within the same school. They found that if two students have a different rank in their own cohort but have the same ability, the student with the higher rank is more likely to finish high school and attend and graduate with a four-year university degree than the student with the lower rank. They identified several mechanisms behind the effect of rank. Students with higher rank have higher expectations for their careers, are more self-confident, and receive more attention from teachers.

4.3.1. Comparing Effort vs Outcomes

Outside of choosing what type of social comparison to use e.g., rank vs mean, another important consideration is to determine what types of activities to socially compare. Previous studies have provided social comparison-based feedback on outcomes such as test scores as outcomes are readily available and a simple measure to compare in a traditional classroom setting. However, a potential reason to compare effort instead of outcomes is that effort is assumed to be tightly correlated with outcomes. If a student puts more effort into studying, then they should achieve a higher score. Several traditional measures of effort have been constructed before, such as number of absences (Elsner & Isphording, 2016), number of homework assignments completed (Natriello & McDill, 1986), and confoundingly, grades themselves (Paredes, 2016). In general, effort as previously measured has been correlated with higher test scores. However, with the advent of Massive Open Online Courses (MOOC) more measures of effort are available. Recent studies have attempted to enable social comparisons of student effort in real time using new

measures of effort such as average time on learning platform per week and time watching videos.

Davis *et al*. (2017), using data from four online MOOCs, compared how student performance varied as a function of providing social comparisons of student effort. Participants were either assigned to a treatment group or a control group. In the treatment group, participants had access to a feedback system that showed how their online activity compared to the behaviour of previous successful students. The control group did not have access to this feedback system. In aggregate, the completion rate for the treatment group was 3.4 per cent higher than the control group. Furthermore, participants with higher education levels benefitted more from access to the feedback systems than those with lower levels of education.

4.4. Social Comparison

Rank-based educational interventions have been used in an attempt to help students plan course selections via a randomised encouragement design. Study participants were either assigned to a treatment group or a control group. In the treatment group, participants were encouraged to use a course planning tool. The tool showed the performance distribution and the intensity distribution (number of hours worked per week) for each class by students who took each class. The control group received no encouragement but could still access the planning tool. Overall, encouraging use of the tool lowered students' grade point averages (Chaturapruek *et al.*, 2018).

Another study examined the role of exposure to peer excellence. It utilised data from a MOOC that had 5,740 students who read and assessed at least three of their peers' essays. There was a negative linear relationship between essay-portfolio quality reviewed by a student and the probability of a student completing the MOOC. A one standard deviation increase in essay portfolio reviewed by a student, reduced that student's probability of completing the MOOC. Moreover, they found that the magnitude of the effect was quite substantial. Assessing the highest quality

essays had an equal to or greater effect than writing the highest quality essays relative to writing an average essay (Rogers & Feller, 2016).

Social comparison has also been used in more traditional classroom-based settings. Azmat, Bagues, Cabrales, and Iriberri (2019) examined how rank and prior beliefs affected students' academic performance. Students' prior beliefs were elicited and in general students were unaware of their rank in the grade distribution and underestimated their own position in the distribution of grades. Students were assigned either to a treatment or control group. The students in the treatment group were shown their rank in distribution every six months over three years. Overall, there was a heterogeneous effect of rank on student performance. The negative effects of rank on performance were largely driven by students who underestimated their rank position due to not receiving feedback. Students who initially underestimated their scores responded positively to the feedback. Another interesting finding was that the effect of ranking was limited to the first time students received their rank (Azmat *et al.*, 2019). In an earlier study, Azmat and Iriberri (2010) provided mean-based feedback and found a subsequent five per cent increase in students' grades. Unlike in their later study, they found a homogenous effect of the mean treatment.

In another study using only absolute feedback, Bandiera, Larcinese, and Rasul (2015) find that absolute feedback (as opposed to no feedback whatsoever) improved students' future exam scores by 13 per cent of a standard deviation. The effect of rank treatments has also been tested (sometimes inadvertently) on teachers as well as students. The Los Angeles Times in 2010 publicly released value added ratings, a metric used to measure the performance of teachers. After the ratings were publicly disclosed, students in low rated teacher classrooms improved their scores in math and English. It was estimated that the magnitude of the rank effect was equivalent to having a 1.2 standard deviation increase in the average value added score of peer teachers (Pope, 2019).

4.5. Small Stakes Feedback vs Large Stakes Feedback

One moderator of how students might respond to social comparison is the stakes of the test. If a test is worth 50 per cent of their grade and a student receives a bad score, the student's subsequent effort should increase on the next test relative to a less consequential test. What is less clear is how students react to feedback on small stakes assessments that precede larger stakes assessments. In the present study, students are given a series of small stakes quizzes before a large stakes exam. Importantly, the content of the small stakes quiz is also covered on the exam. One way students could react to how they scored on the scores quizzes is to direct their effort proportionally to the weight of the exam. Another possibility is that they take their score on the quiz as a signal of how they will perform on the final exam, which determines the majority of their score. Subsequently, they either increase or decrease how much (or less) effort they need to put into the class based on their score on the quizzes.

Whether social comparisons increase effort on small stakes quizzes is an open question. There is evidence to suggest that students respond strongly even to absolute grading scales. Main and Ost (2014) find that students who score just below an 80, which is the cut-off for a B in many grading scales, on the first exam are more likely to achieve higher marks on the second exam than students who scored slightly above an 80 on the first exam. With students responding so strongly to even absolute scales, students might respond even more strongly to social comparisons.

4.6. Reminders

Our study can be viewed as a study on both feedback and reminders. The present study centres around reminding students of feedback that they had previously received. Using reminders in the domain education is commonplace. Reminders in the form of text messages have been used to ensure that students in high school complete the requisite tasks to matriculate into college (Castleman & Page, 2015) and increase the attendance rate for adult learners as well (Chande *et al.*, 2015).

Reminders are also often sent to parents of students. In one study, children whose parents were opted into a text messaging service that informed them of their child's academic performance and attendance saw a 0.05-0.06 point increase in their GPA (Bergman & Rogers, 2017). Another study utilising parental reminders found that sending weekly automated reports to parents of middle and high school students increased class attendance by 12 per cent and reduced course failures by 28 per cent (Bergman & Chan, 2019). Personalising text messages have also been shown to be effective at the Pre-K level. In another study parents who received personalised and differentiated texts children were 50 per cent more likely to move up a reading level relative to their peers who did not receive personalised text messages (Doss *et al.*, 2017).

4.7. Predictions

In the present study we test the effectiveness of various social comparison-based reminders. We use the social based comparison used in Taylor *et al*. (2015) that pits the effectiveness of absolute vs mean vs rank-based feedback. They found that rank based comparisons were the most effective. Accordingly, our predictions are as follows:

1A. Rank-based reminders will be more effective at increasing quiz scores compared to mean-based reminders.

1B. Rank-based reminders will be more effective at increasing information seeking than mean- based reminders.

2A. Mean-based reminders will be more effective at increasing quiz scores compared to absolute-based (score only)[2] *reminders.*

[2] Absolute-based reminders are comprised of the score only and no comparisons.

2B. Mean-based reminders will be more effective at increasing information seeking compared to absolute-based reminders.

4.8. Experimental Design and Data

4.8.1. Background

The experiment took place at a mid-size university in the West Midlands, UK. The class was an operations course, which consisted of 420 students and was spaced out over 10 weeks. The class followed an inverted format where students were expected to watch online videos each week and then are expected to take an in-class quiz on the content of the videos. In total, there were four quizzes one for each thematic block

4.8.2. Experimental Design

In each week of the module, a thematic sequence was covered. At the end of each block, students must complete a quiz on the material, which was covered in that block. Each week a group was randomly assigned to one of three feedbacks treatments. The order in which each group received the treatment is shown in Table 4. This treatment was in addition to the feedback that they normally receive. Every student was randomly assigned to one of three groups:

1) **Absolute**: Students receive an email reminding them of the score they received on the previous quiz.
2) **Mean**: Students receive an email reminding them of how their score compared to the mean score on the previous quiz.
3) **Rank**: Students receive an email with their rank on the last quiz.

After each quiz students were emailed their scores in the form of one of the treatments. Additionally, a link to a webpage which included more information about their scores was included in all of the emails.

	GROUP 1	GROUP 2	GROUP 3
BLOCK 2	Rank	Mean	Absolute
BLOCK 3	Mean	Absolute	Rank
BLOCK 4	Absolute	Rank	Mean

Table 4: The Treatment Each Group Received by Block

4.8.3. Data Collection

The analyses in this study used commonly collected administrative data including their score from in-class quizzes. Moreover, data was also collected on how many times students' viewed additional information about their scores. The scoring of each quiz was automated.

4.8.4. Data Analysis

There is one primary outcome and one secondary outcome. The primary outcome of interest is how quiz scores vary as a function of the different reminders. The secondary outcome of interest is the effect of treatment on information seeking as measured by webpage views. After each thematic block, students were given a quiz; the outcome variable is the score students received on each quiz. The treatment is the type of reminder they received on the quiz in the preceding week.

A linear mixed model with by-participant random intercepts was used to estimate the effect of treatment on quiz scores. *P* values were constructed using the Kenward-Roger approximation for degrees of freedom as implemented in the afex package (Singmann *et al.*, 2016). A generalised linear mixed model, with binomial family and logistic link function, with by-participant random intercepts was used to estimate the effect of treatment on views. *P* values for the generalised linear model were determined using Likelihood Ratio Tests as implemented in the afex package.

4.8.5. Attrition

There were 420 students in the class. Seven students were excluded from the analysis because the email addressed them by the wrong first name. Additionally, four students were not assigned to a treatment group. In total, this left 409 students in the sample. For each student, only their top three (out of four) quiz scores counted towards their grade. Forty-nine students missed at least one quiz. Table 5 shows the number of quizzes missed each block.

	BLOCK 1	**BLOCK 2**	**BLOCK 3**	**BLOCK 4**
MISSED QUIZZES	13	2	9	28

Table 5: Missed Quizzes by Block

4.9. Results

4.9.1. Quiz Scores

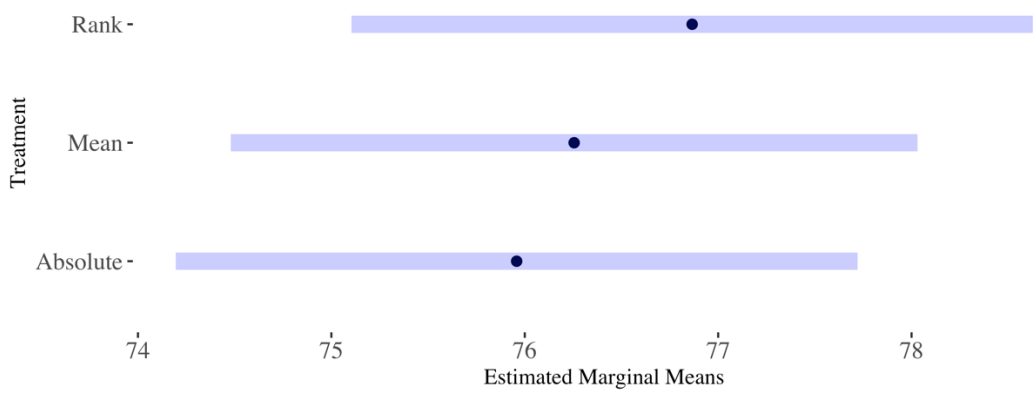

Figure 24:Estimated Marginal Means for each treatment type with 95 per cent confidence intervals.

Figure 24 shows the results of treatment over the blocks. The results show that students who received the rank treatment in the preceding block increased their quiz scores in the subsequent block by 0.9 points relative to the absolute treatment. Students who received the mean treatment in the preceding block increased their

quiz scores in the subsequent block by 0.3 points relative to the absolute feedback treatment. Overall there was no main effect of treatment $F(2, 792.03) = .29$, $p = .75$.

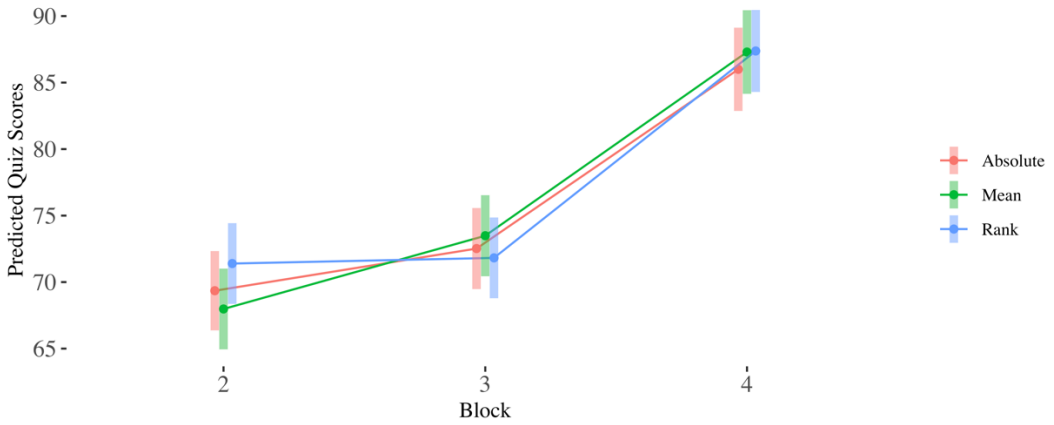

Figure 25: Predicted quiz scores for each treatment over blocks. The bars indicated 95 per cent confidence intervals.

Figure 25 shows the predicted means by block and the treatment received based on the previous block's quiz score. For the rank condition, the difference in estimated marginal means appears to be driven in large part in week two.

Additionally, there appears to be large differences in the difficulty between quizzes given the large differences between estimates block over block $F(2,791.81) = 112.18$, $p < .0001$. The quiz scores for block four differed significantly from block three $t(798) = -14.28$, $p < .0001$ and block 2 $t(795) = -14.108$, $p < .0001$, but not for blocks two and three $t(783) = -2.506$, $p < .0001$. However, the effect of treatment does not appear to vary over block $F(4,794.19) = .77$, $p = .54$. Another possible interpretation is that rather than the quizzes being more difficult, students applied more effort to different quizzes.

4.9.2. Information Seeking (Views)

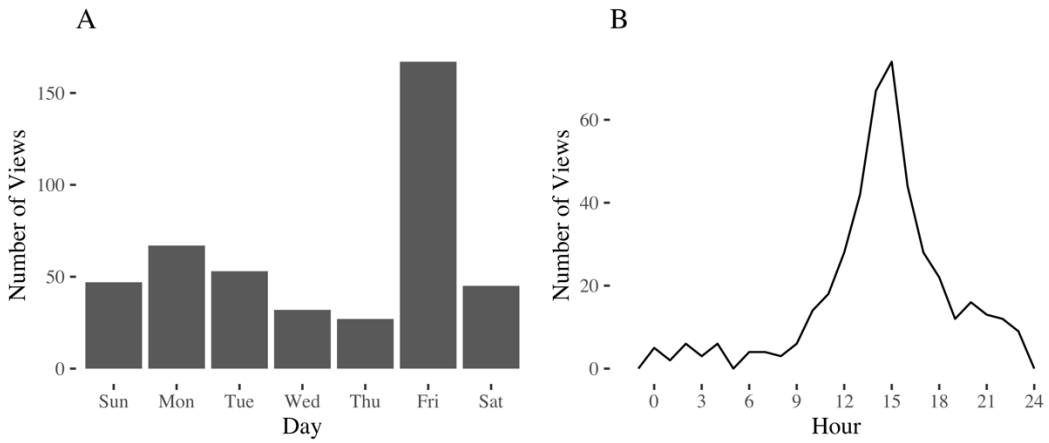

Figure 26: Panel A shows the number of page view on each day. Panel B shows the number of page views for each hour of the day.

4.9.2.1. Number of views by day and hour

One potential issue with our design is that we have no direct way to attribute page views to our intervention. However, emails after the block one and two quizzes were sent out on a Friday and the email after block three was sent out on a Monday. All three emails were sent in the afternoon. Figure 26A shows that most views on any given day of the week are on Friday and Monday. Also, Figure 26B shows that page views happen most frequently in the afternoon. Taken together, this provides some evidence that emails encourage students to view more additional information about their marks. This additional information consisted of a histogram of their marks.

4.9.2.2. Views by intervention and interval

The secondary outcome of interest was encouraging information seeking (number of views). Specifically, we directed students to a link that led them to a webpage which contained additional information about their scores.

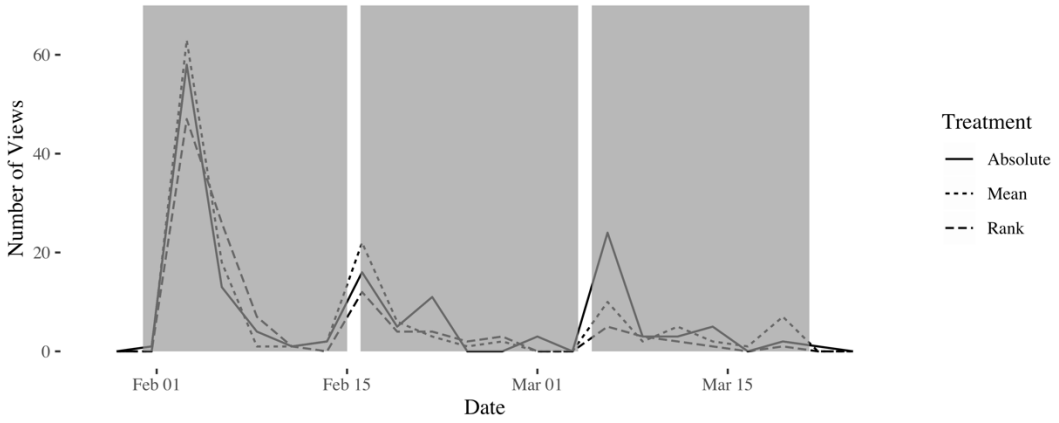

Figure 27: Number of views overtime by treatment. The shaded areas represent the three different intervals. The dotted lines represent the treatment groups.

The first two intervals were defined as the period between reminders, 16 days for the first interval and 17 days for the second interval. The first interval was 16 days and the second interval, 17 days. The third interval consisted of 16 days after the quiz in block four. Figure 27 shows the number of views over time by each treatment type for every interval. In general, the number of views in total is fairly low over every interval. Interval one clearly had the most views, while intervals two and three had a similar number of views. Unsurprisingly, views spike after the beginning of the interval, which is close in proximity to the first day they are able to view their scores.

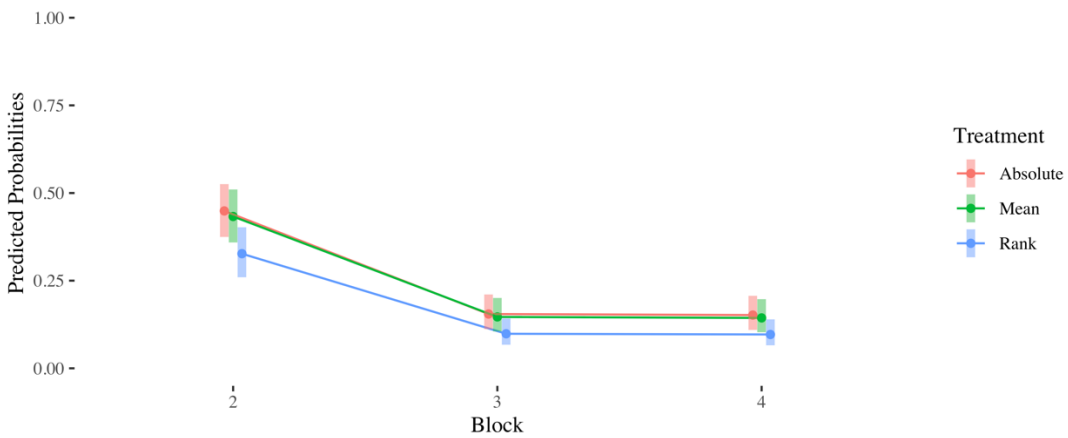

Figure 28: Predicted probability of a page view for each treatment by block. The bars indicate 95 per cent confidence intervals.

Overall, there was a significant effect of treatment $\chi^2 = 9.15$, $p = .01$. Participants who received the mean treatment were equally likely to view additional information about their scores as participants who received the absolute treatment; the magnitude of the effect (odds ratio) was negligible OR = .94 [95% CI: 0.62,1.43][3]. However, participants who received the rank treatment were less likely to view additional information about their scores OR = .94 [95% CI: 0.62,1.43] than when receiving the absolute treatment. Likewise, participants who received the rank treatment were less likely to view additional information about their scores OR = .64 [95% CI: 0.41,1] than when receiving the mean treatment (Figure 28). There was also a main effect of interval $\chi^2 = 100.20$, $p < .0001$. Participants were more likely to view additional information about their scores during interval 1 than interval 2 OR = 4.44 [95% CI: 2.85,6.99] or interval 3 OR = 4.55 [95% CI: 2.89,7.17]. Moreover, participants were not more likely to view additional information about their scores in interval 2 than in interval 3 OR = 1.02 [95% CI:.64,1.65].

4.10. Discussion

In this experiment we tested how different forms of feedback influence how students choose to allocate their effort. In line with our predictions and the literature on descriptive norms, the rank treatment improved quiz scores the most, followed by the mean, and absolute treatment. However, our results should be viewed with some scepticism given the width of the confidence intervals and lack of statistical significance.

For the secondary outcome (number of views), we observe the reverse pattern results than with the primary outcomes. The absolute treatment was the most likely to encourage information seeking followed by the mean, and rank treatments. This is also in contrast to the findings in (Taylor *et al.*, 2015), which find that the rank treatment is the most effective treatment for increasing information seeking. A likely reason for this is that students wanted more information about their scores, specifically where they rank. Our study deviates from Taylor *et al*. (2015) in several

[3] Odds ratio is abbreviated as OR.

different ways, which may explain why we see a reversal in the effectiveness of the treatments for information seeking. Firstly, excessive drinkers are unlikely to readily have access to information about excessive drinking given they do not know they are excessive drinkers in the first place. In our context, participants certainly are aware they have taken a quiz and are more likely not as "surprised" about being ranked. Another interpretation of our results, more in line with our initial predictions, is that after seeing their rank participants felt that they no longer needed any additional information about their scores. However, the mean and absolute information initially communicated still left them with the desire for more information. Moreover, the outcomes our intervention seeks to influence are considerably lower stakes than in Taylor *et al.* (2015).

The most unambiguous use of rank to improve outcomes has been in the case of health behaviours such as in Taylor *et al.* (2015). One reason why rank might be more effective in the domain of health is that there is typically limited direct social comparison. For example, people do not get together in an office and compare how many units of alcohol they drank each week and then determine their rank. Thus, people might react more strongly to receiving feedback on behaviours about which they do not usually receive feedback. In the domain of education there is a deluge of feedback, especially so in the current module. This may have caused students to pay less attention to a feedback reminder compared to all of the other feedback that they usually receive.

One potential reason we observe a mild effect of treatment (although the pattern of results was the opposite of our predictions) on information seeking is that participants simply have to click a link, which is a one-shot action. If we were to observe an effect of treatment on quiz scores, this would imply a more sustained behaviour change week over week e.g., participants put more effort into studying. The lack of sustained behaviour change due to a behavioural intervention has also been noted in other domain such as exercise, weight loss, and charitable giving (Brandon *et al.*, 2017). In our study, the declining efficacy of our intervention is best illustrated in Figure 28, where the number of webpage views declines over time regardless of treatment group.

4.11. Limitations and Future Studies

There are some limitations to be noted in this study. Firstly, students could access a histogram comprised of all the quiz scores at any time. This could have reduced any potential differences we could have observed between the social comparison reminders (mean and rank) and the absolute condition. Secondly, after each quiz the software used for the quizzes sends each student the mean and range of the previous quiz. Due to reasons for ethics approval, we were only able to add additional information, not remove information.

In regard to the emails, we were not able to directly attribute whether students viewed more information about their quiz scores because of our intervention. Although, given that the number of views spiked on the days and hours closely after we sent out the emails, we can reasonably attribute some of the views to our intervention.

Previous research suggests that students' ability plays a large role in how they respond to social comparison-based feedback. Using the data in our study we found that initial ability plays a strong role in how students perform in general. Moreover, having an exogenous baseline measure of ability would further support the above finding while additionally enabling us to disentangle the effects of the social comparison reminders on higher and lower ability students.

Another limitation of our study is that we do not know the "surprisingness" of our treatment. That is, what participants prior beliefs about their rank were. Future studies should more tightly control the information students have access to regarding their rank in the classroom. Additionally, eliciting prior beliefs of participants would also be a useful endeavour. As shown in previous studies, the effect of rank is greatest when participants beliefs conflict with reality.

Spillover effects between students may have also occurred. Students may have compared emails with each other. Moreover, every student received each

treatment, which may have increased the possibility of reminder fatigue. Future studies should further account for these spillover effects.

Due to the online nature of the module, there was opportunity to understand how different forms of social comparison reminders encourage interaction with the online materials. Unfortunately, the infrastructure to collect this data was not in place while the trial was being conducted. Future studies should investigate how social comparison reminders influence interaction with online resources. Another, important variable to take into account is students' baseline ability. Students with higher and lower ability may react differently to social comparison reminders. We attempted to collect this data but were unsuccessful due to data protection stipulations.

One advantage of randomised control trials is the increased external validity. However, this increased external validity can comes at a price decreased internal validity. When using lab based studies for examining intertemporal choices the trade-offs are known because they are largely determined by the researcher. However, in this study not all of trade-offs are known as students could be trading off time between many things that occur in their daily life.

The present study focused on providing social comparison reminders which compared outcomes (quiz scores). Past measures of effort have used outcomes as a measure of effort, yet outcomes are not a direct observation of students' actual effort (studying behaviour). Inverted classrooms offer a new and more direct opportunity to observe student effort as much of the course content is hosted online. Instead of using social comparisons on outcomes, future studies could use social comparisons of effort.

Chapter 5. Conclusion

5.1. Summary

The research presented in this thesis investigated the role of learning and feedback in intertemporal choice and in education. This dual track of investigation has resulted in several notable contributions using lab and field-based studies. To investigate the role of experience in intertemporal choice, a novel methodological paradigm was employed using ideas from both the risky and intertemporal choice literature. To investigate the optimal form of feedback, I conducted a randomised control trial where the efficacy of different feedback was tested.

In Chapter 1, the many facets of intertemporal choice are reviewed. The normative model of intertemporal choice, the discounted utility model, is contrasted with what is now the dominant model of intertemporal choice, hyperbolic discounting. Next, the common difference effect and the influence of time perception on intertemporal choice are reviewed. The next section of Chapter 1 reviews a subdomain of risky choice literature, the decisions from experience literature, which — until recently — was a separate literature. Important findings from the decisions from literature are discussed such as when decisions from description diverge from decisions from experience, often referred to as the description-experience gap. Crucially, towards the end of Chapter 1, I identify how past experiments have incorporated how experience and why these experiments are not designed to adequately address whether description based intertemporal choices differ from experience based intertemporal choices. Next, I discuss similarities between risk and time, specifically the common difference and common ratio effects and how experience might diminish the size of these effects. Lastly, I review the most recent intertemporal choice from experience experiments and highlight how different delays are used in description and experience.

In Chapter 2, the pair of intertemporal choice experiments yielded several novel contributions. We investigated whether a description-experience gap existed in intertemporal choice by employing a sampling paradigm imported from decisions

under risk, delays on the order of seconds, and real consumable rewards for the first time in one experiment. Past experiments have also addressed whether a gap in description-experience exists in intertemporal choice. However, there are several areas where the experiments detailed in this thesis offer improvements over previous methodology. For the experiments in Chapter 2, participants in the sampling phase in the experience group actually sample and experience the same delays as those in the decision phase — ensuring description is not mixed in with experience. Sampling paradigms in previous intertemporal choice from experience experiments have often mixed description with experience. People simply sample descriptions (Dai *et al.*, 2019) rather than the experience actual delays themselves. Another improvement we make over previous designs is that the timescales between description and experience are similar; unlike in other experiments where time scales between description and experience have differed (Lukinova *et al*., 2019). To my knowledge, we also employ for the first time a novel reward in intertemporal choice, music.

The methodological contributions in Chapter 2 have also enabled the discovery of several novel empirical results that contribute to the literature of intertemporal choice. Firstly, we asked whether a description-experience gap exists in the domain of intertemporal choice. Secondly, does the common difference effect occur when delays are learned from experience.

In Experiment 1, there was no overall difference in patience between description and experience. Moreover, we found preliminary evidence for a common difference effect in both description and experience. In Experiment 2, the same procedure was used except the delays were closer together and thus harder to distinguish: 5s vs 15s and 85s vs 95s. Again, there was no overall difference in patience between description and experience. Unlike Experiment 1, however, the common difference effect was only evident in description but not experience, suggesting a description-experience gap exists in intertemporal choice.

One potential reason for the selective appearance of the common difference effects is that the uncertainty about the delays in experience may obscure temporal

distinctiveness, especially when the delays to be compared are similar in duration. Thus, in Chapter 3, we aimed to make descriptions more like experience. To do so, we included conditions where the delays were described as point estimates as well as ranges, with a ±40 per cent range, drawn from time estimates in the first two experiments. With this design, there was a common difference effect across both the range and point conditions. Together Experiments 1 and 2 establish that the common difference effect occurs even with very short delays learned from experience. In contrast to the literature on risk, however, we do not find a consistent description-experience gap in intertemporal choice. Finally, temporal uncertainty did not appear to attenuate the common difference effect.

In Chapter 4, I ran a randomised control trial to empirically test different forms of feedback reminders. I developed an interventional design to compare the effect of rank-based vs mean-based vs absolute based feedback on subsequent student performance. Overall, the results suggested that rank based feedback was the most effective feedback followed by mean-based then absolute-based feedback; however, these results were not significant. Interestingly, we found the reverse pattern for our secondary outcome, whether students viewed additional information about their grade, which provides additional support for the hypothesis that people are primarily concerned with their rank. I also investigated the role of initial aptitude on student performance and found that consistent with other skill-based activities, initial aptitude plays an outsized role on student performance. The prominence of initial aptitude suggests an upper limit to different educational interventions of a similar nature to the one described in Chapter 4.

5.2. Limitations

Despite the several notable contributions in this thesis, there are some limitations. In the intertemporal choice experiments in Chapters 2 and 3, delays consisting of seconds instead of delays on the order days, months, and years are used. Short delays are used because of the logistical and financial requirements of keeping participants in the lab for longer periods of time. One could imagine an experiment with much longer delays using hours instead of seconds. Another

limitation to take into consideration is whether the results from the experiments using seconds generalise to much longer delays that are commonly found in the literature. Outside of the larger issue of whether the results scale to longer delays, only the delays shown in Table 6 have been used. Moreover, the delays were fixed and were not titrated as often found in other tasks (Bickel, Odum & Madden, 1999).

Experiment	**Groups**	**Delays**
Chapter 2:	Description Early	5 vs 25
Experiment 1	Description Late	45 vs 65
	Experience Early	5 vs 25
	Experience Late	45 vs 65
Chapter 2:	Description Early	5 vs 15
Experiment 2	Description Late	85 vs 95
	Experience Early	5 vs 15
	Experience Late	85 vs 95
Chapter 3	Points Early Narrow	10 vs 20
	Points Early Wide	5 vs 25
	Points Late Narrow	90 vs 100
	Points Late Wide	85 vs 105
	Range Early Narrow	6-14 vs 12-28
	Range Early Wide	3-8 vs 15-35
	Range Late Narrow	54-126 vs 60-140
	Range Late Wide	51-119 vs 63-147

Table 6: Trial Types across Chapters 2 and 3

As stated above, the vast majority of intertemporal choice studies use money (both real and hypothetical) as rewards, while the intertemporal choice studies in this thesis use music as rewards. A limitation to this approach is that we are not able to fit delay discounting models to the data because we do not know the precise valuations of each playlist, only that one playlist is preferred over the other.

There are also some further limitations to be noted for the feedback randomised control trial. One important moderator that was unaddressed is baseline ability. While a measure of initial aptitude was constructed, this measure is endogenous given that it is constructed from the data we manipulated. I attempted to obtain exogenous data on students' baselines ability (e.g., GCSE). This effort was unsuccessful, however, despite numerous requests over several years. Consideration should also be given to the population characteristics. This randomised control trial was conducted at a highly ranked international business school. Thus, the characteristics of the sample may influence the generalisability of our results. For example, a rank intervention may be more effective at universities with less competitive admission requirements.

5.3. Implications and Applications

There are several implications and applications for our findings from the intertemporal choice studies. People need to make decisions over time in everyday life — often using their own experiences rather than explicit description. The vast majority of intertemporal choice studies are paper based studies that do not incorporate how people learn from experience. Based on these studies, behavioural scientists make policy and product recommendations for both governments and large corporations that affect millions of people's lives. If a large description-experience gap were to exist in intertemporal choice, this would call into question the generalisability of many previous findings in intertemporal choice and the many recommendations behavioural scientists make based on these finding. Reassuringly, we do not find a description-experience gap in intertemporal choice and find some evidence that different intertemporal choice phenomenon found in description can also be found in experience.

With the advent of technology, people increasingly need not rely on experience; they can simply check their phones for many intertemporal choices such as choosing directions. However, checking can come at a cost (Roth, Wänke & Erev, 2016). On the extreme end of costs, 3,166 people died in 2017 from distracted driving according to the NHTSA (2017). On the more routine end of the spectrum,

costs such as checking one's phone distracts from what might be more meaningful tasks (Jersild, 1927; Hyman *et al.*, 2009). Presumably, people check their phones to improve their intertemporal choices and consequently incur the costs of switching such as decreased performance on the task at hand e.g., driving. Our studies suggest that one need not make this trade-off between distraction and accuracy. Simply relying on past experiences, at least for short scale decisions, may suffice.

There are a number of insights and practical applications that were learned from carrying out the feedback randomised control trial. Firstly, implementing the randomised control trial required considerable effort over several years. It involved building trust with the educational institution as well as demonstrating the potential benefits of experimentation.

Secondly, a common them that was encountered was whether the randomised control trial was "fair". Specifically, whether the treatments would affect each group equally. Accordingly, the RCT was redesigned so that each group received each treatment.

One interesting observation was that this idea of fairness was only being applied to experimentation rather than the status quo. There is no way to understand what is fair if the different effects of feedback are not understood in the first place. Consequently, an important point to emphasise is that a priori we do not know which type of feedback works best and for whom without evidence. Conducting randomised control trials is one step towards accumulating such evidence.

The results from the randomised control trial further emphasise the need to conduct more experiments. The literature suggests that rank is by far and away the most effective treatment, yet our results suggest that social comparison-based feedback may be most impactful in domains where people's activities are typically not ranked. Understanding how people respond to different forms of feedback is important as the advance of the Internet of Things (IoT) and computer vision technology allows further quantification of human behaviour. For example, at the American franchise, Outback Steakhouse, computer vision technology can be used

to tell whether a diner's drink is empty (Matsakis, 2019). An overly enthusiastic manager might be tempted to provide rank-based feedback to their employees on cup refill time.

Perhaps a larger question to ask is do people actually know how they compare to their peers and is the metric being used to compare them to their peers matter to the overall outcome. It is unlikely that water refills are the determining factor as to whether the customers are satisfied with their experience, particularly so outside of the US context. Moreover, it is possible that over quantification by providing social comparison-based feedback could lead to high employee turnover — costing businesses more than the efficiencies they gained through creating metrics of employee behaviour.

Taken together, the implications and applications of the findings in this thesis present a cautionary note on the increasing prevalence of metrics and quantification of human behaviour. As a result of the spread of technologies that are able to measure human behaviour, more metrics are created. However, more metrics may simply result in more problems whether it is because the metrics actually do not actually change behaviour, or they change behaviour in undesirable ways.

Appendix 1

Experiment 1

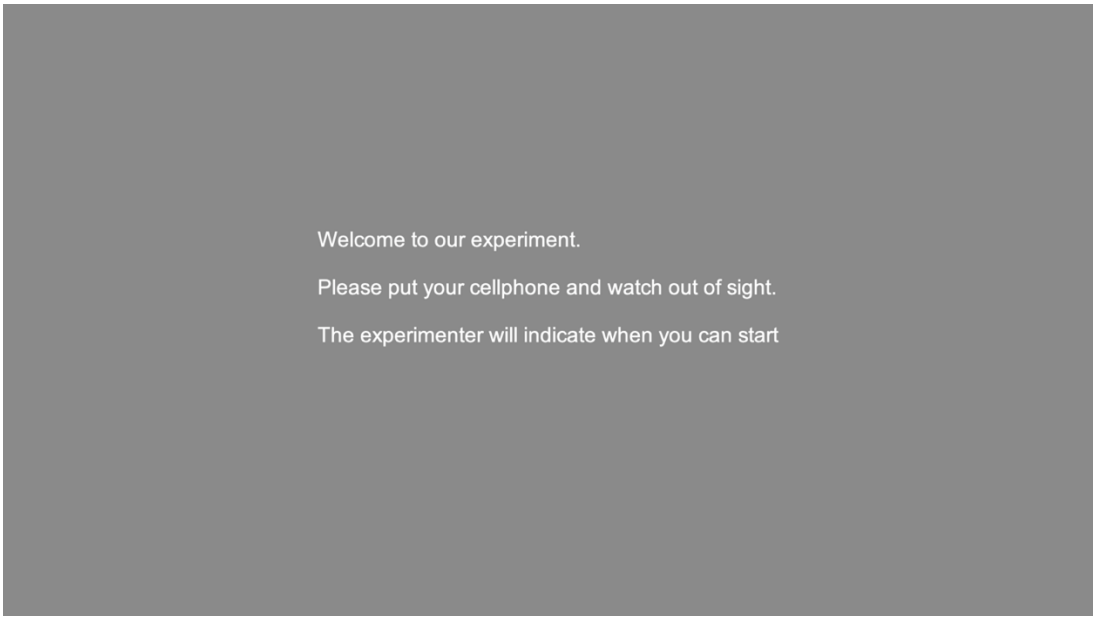

Figure 29: Opening phase

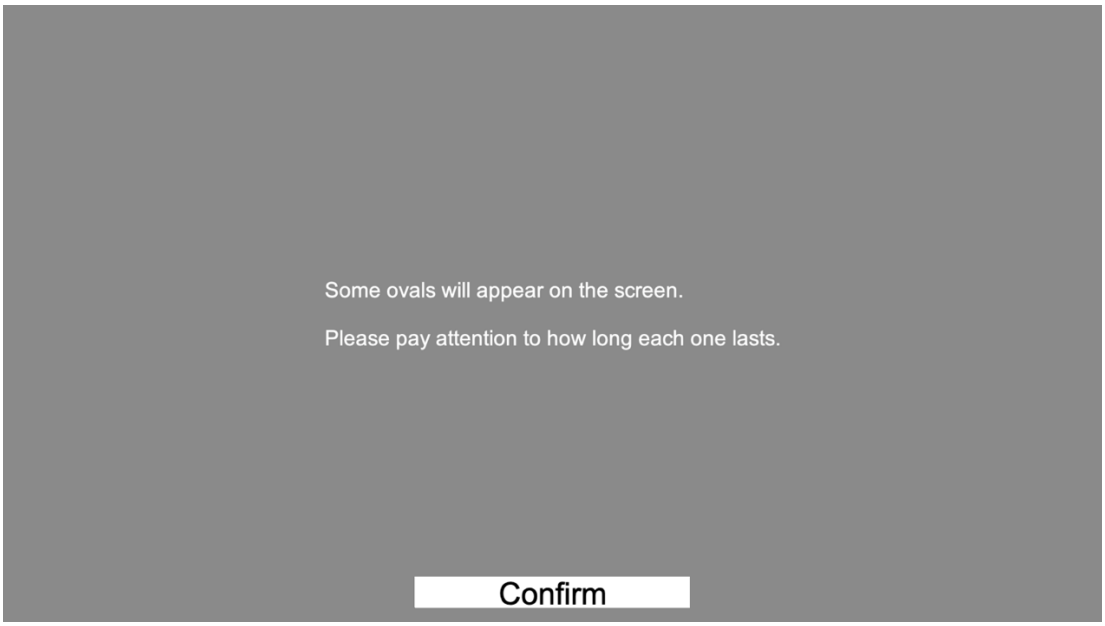

Figure 30: Sampling Phase

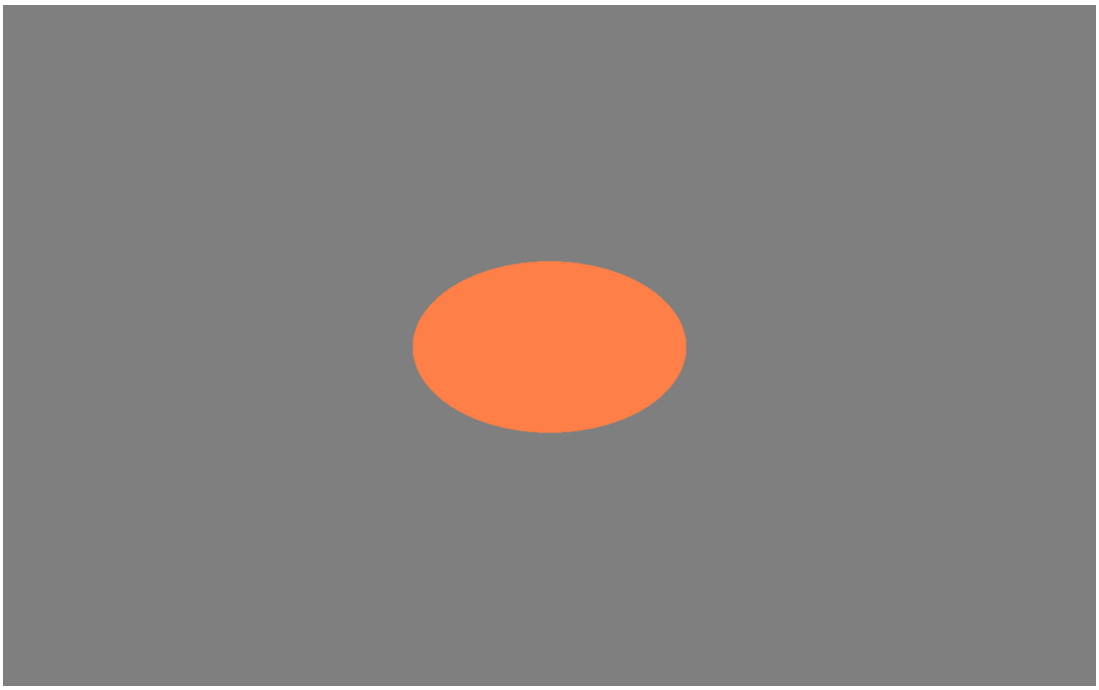

Figure 31: Sampling Phase

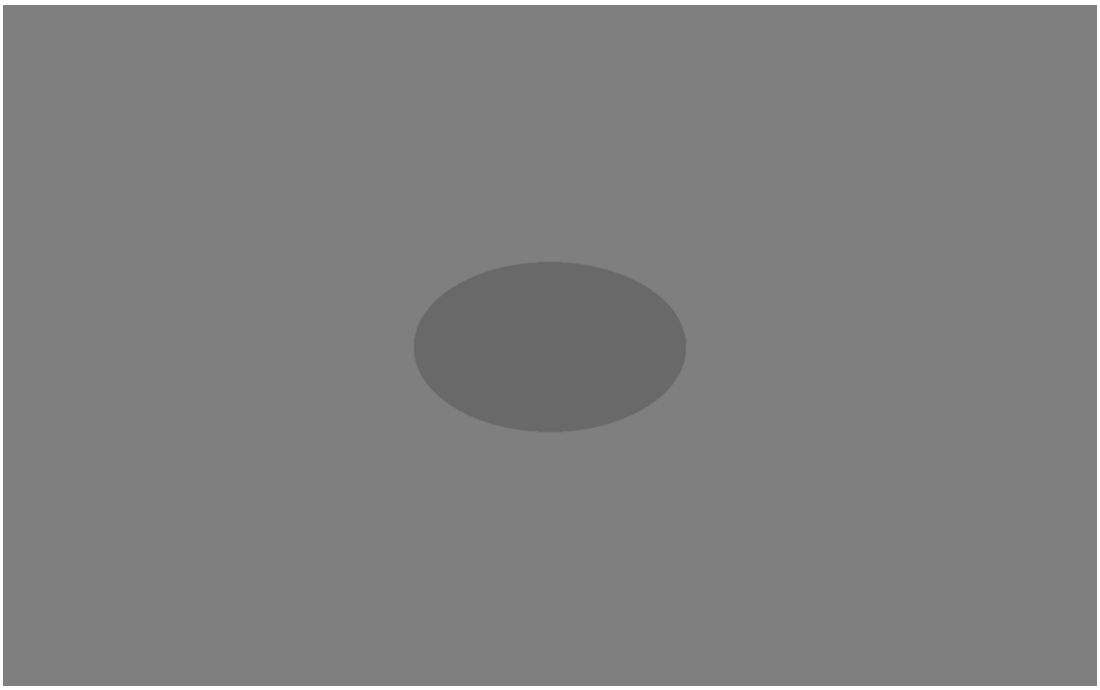

Figure 32: Sampling Phase – Description

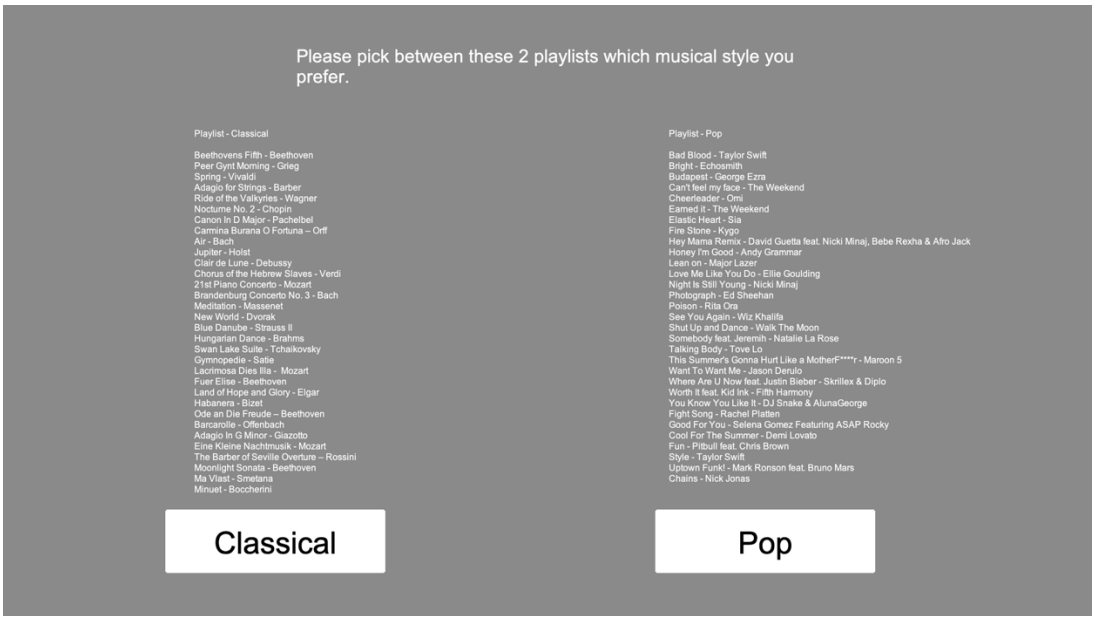

Figure 33: Reward Selection

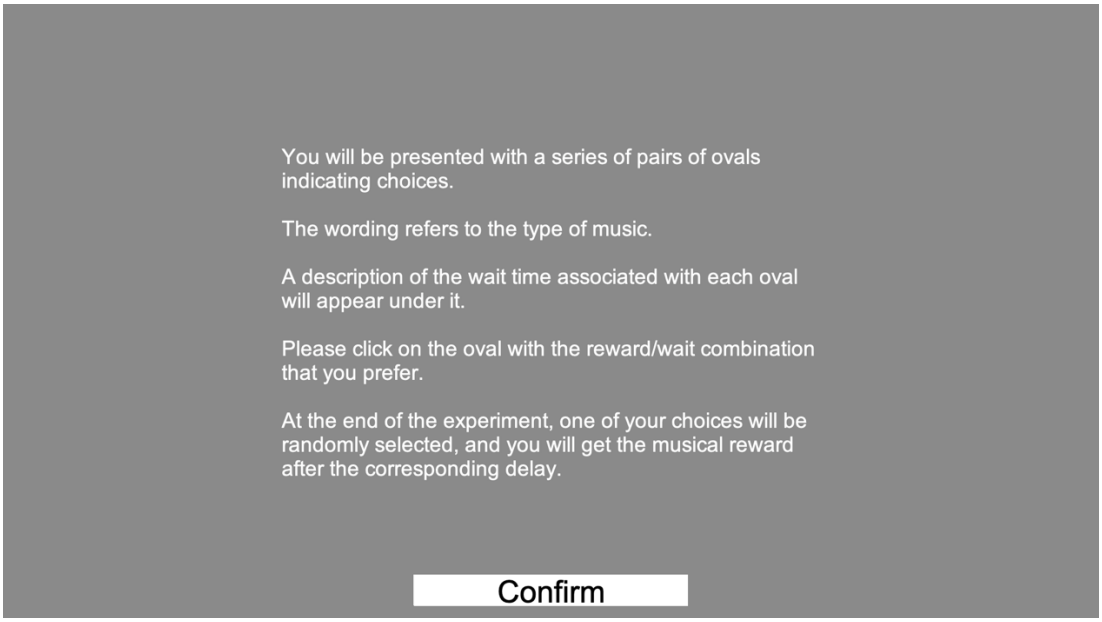

Figure 34: Decision Phase – Description

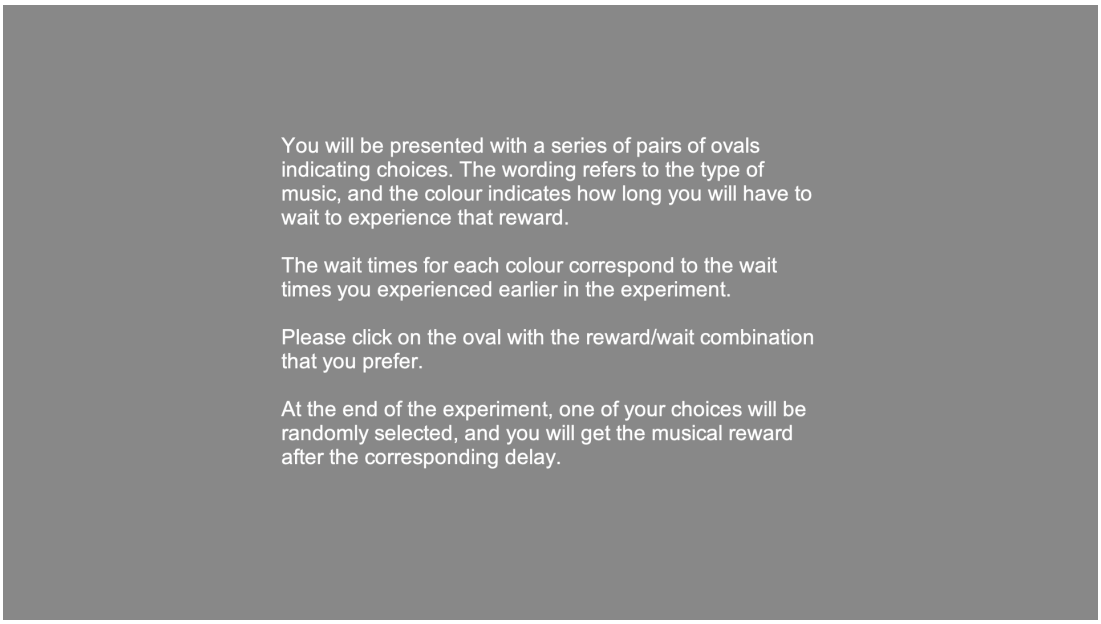

Figure 35: Decision Phase – Experience

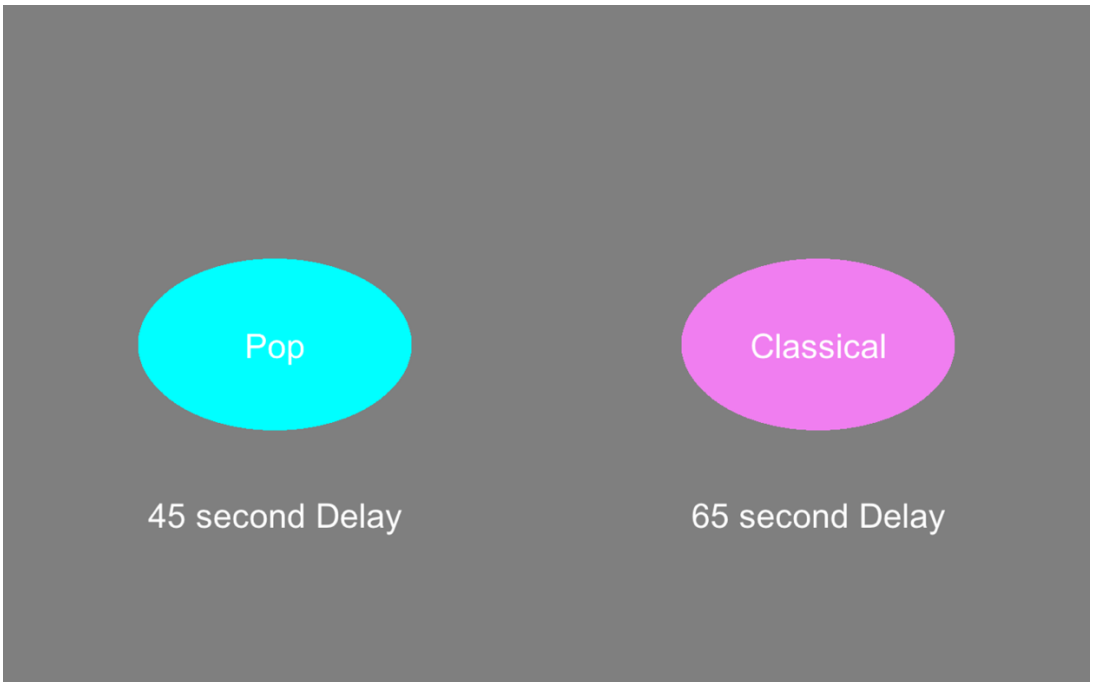

Figure 36: Decision Phase – Description

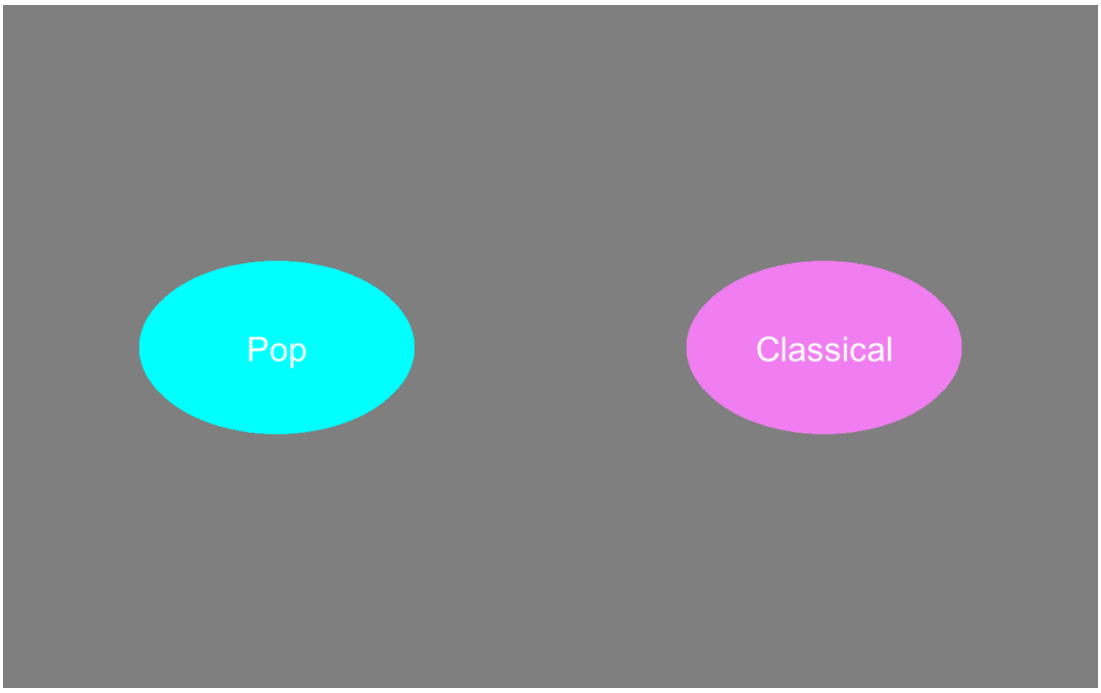

Figure 37: Decision Phase – Experience

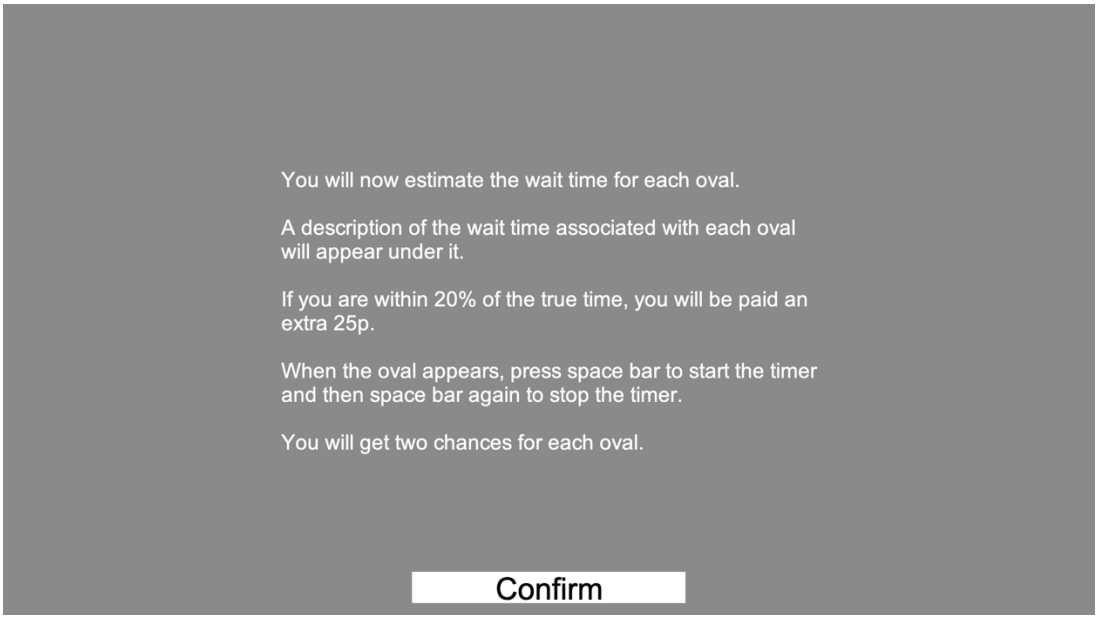

Figure 38: Time Estimation Phase – Description

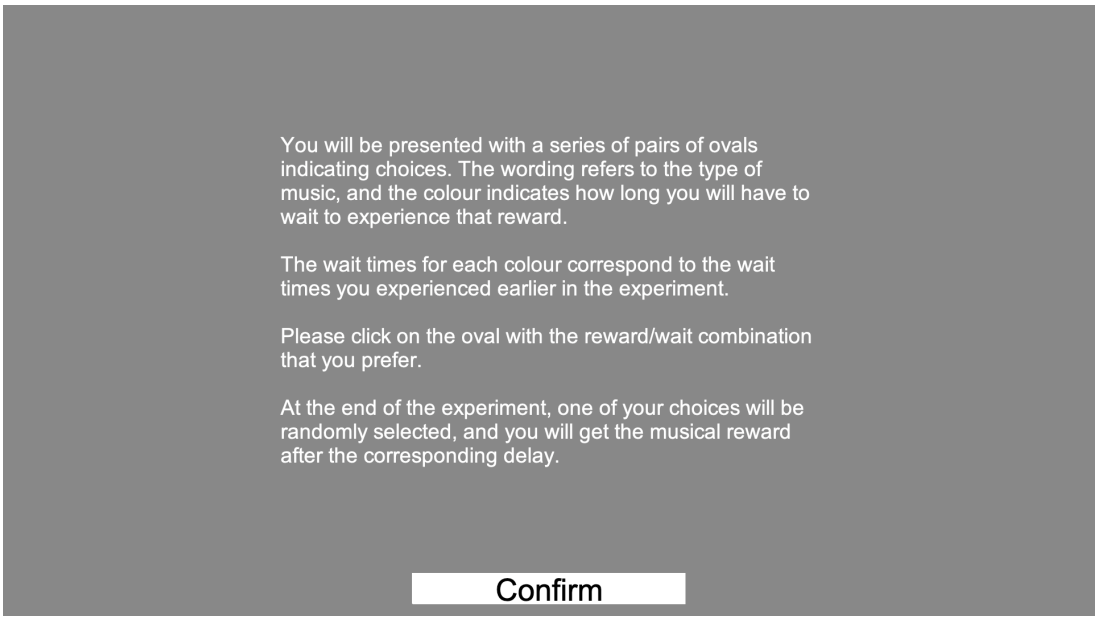

Figure 39: Time Estimation Phase – Experience

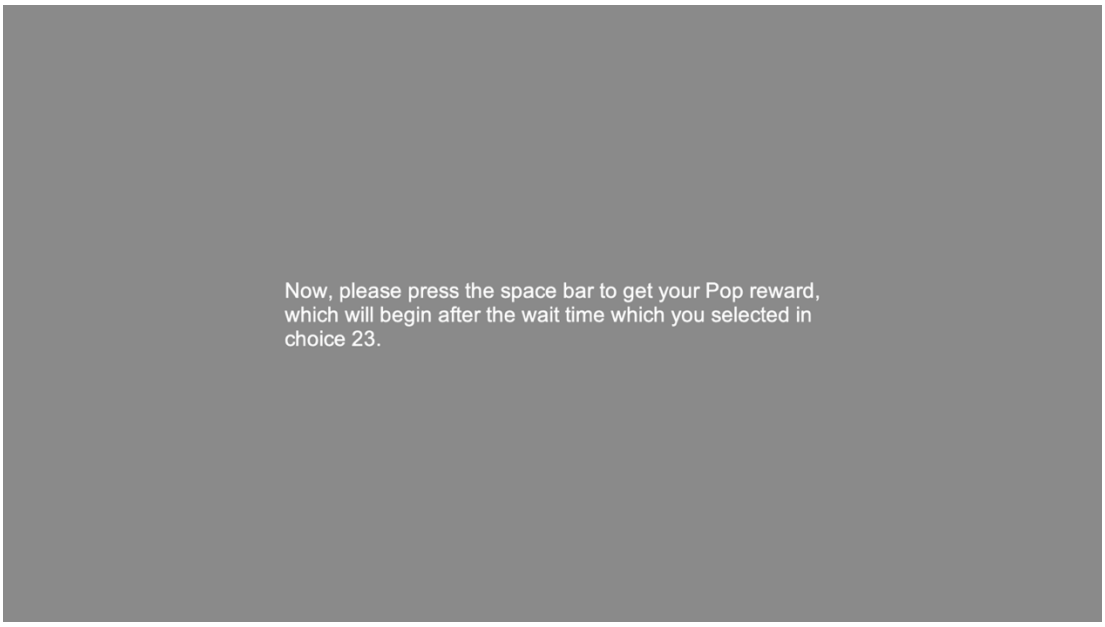

Figure 40: Reward Phase

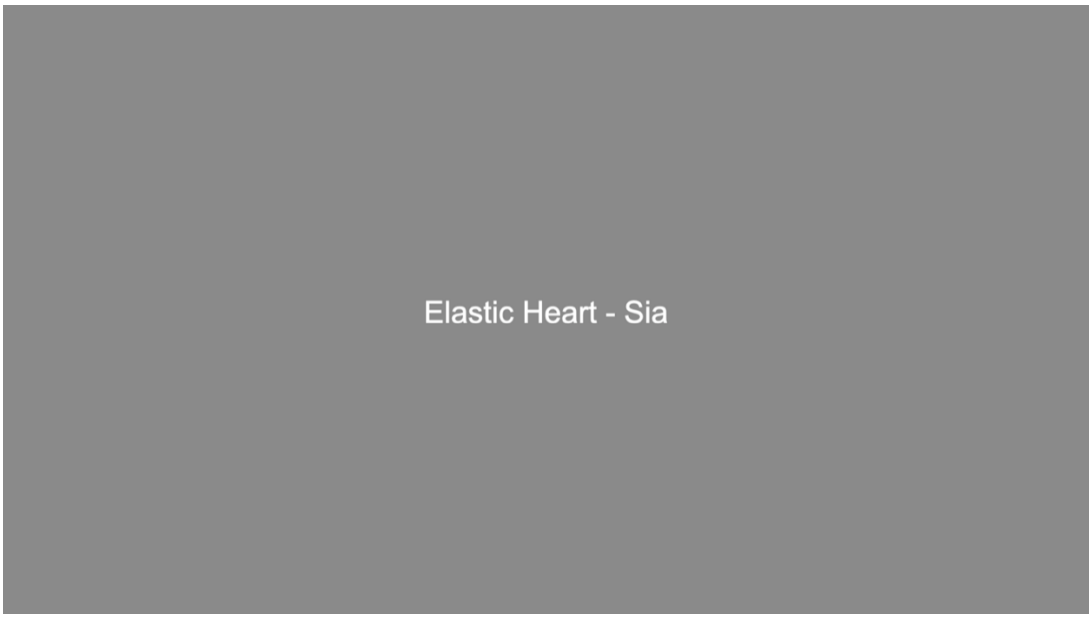

Figure 41: Reward Phase

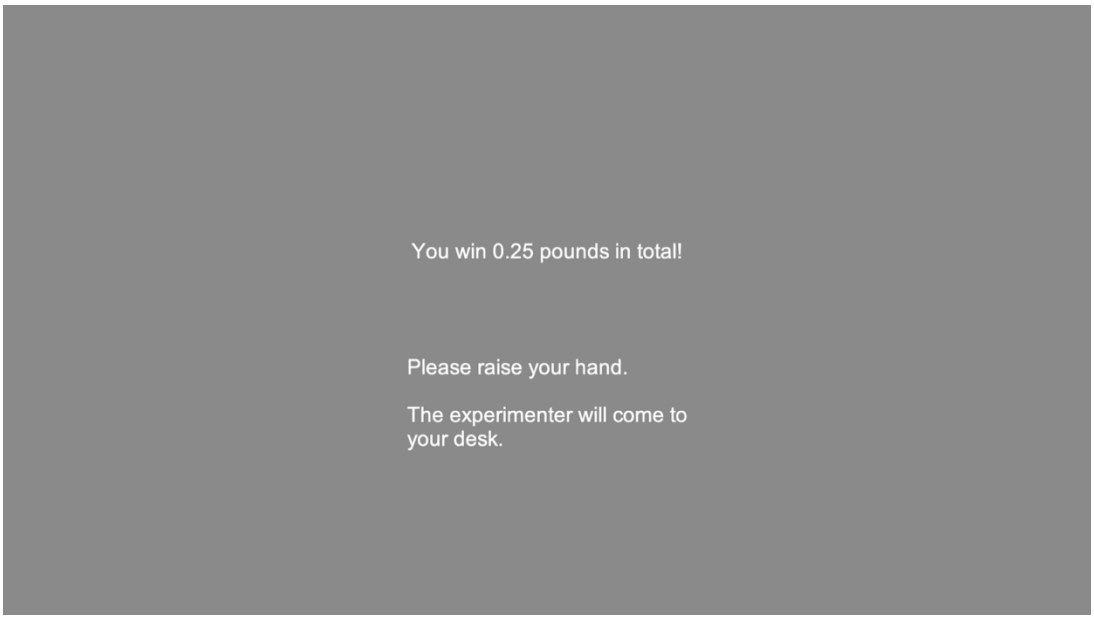

Figure 42: Reward Phase

Experiment 2

Opening Phase

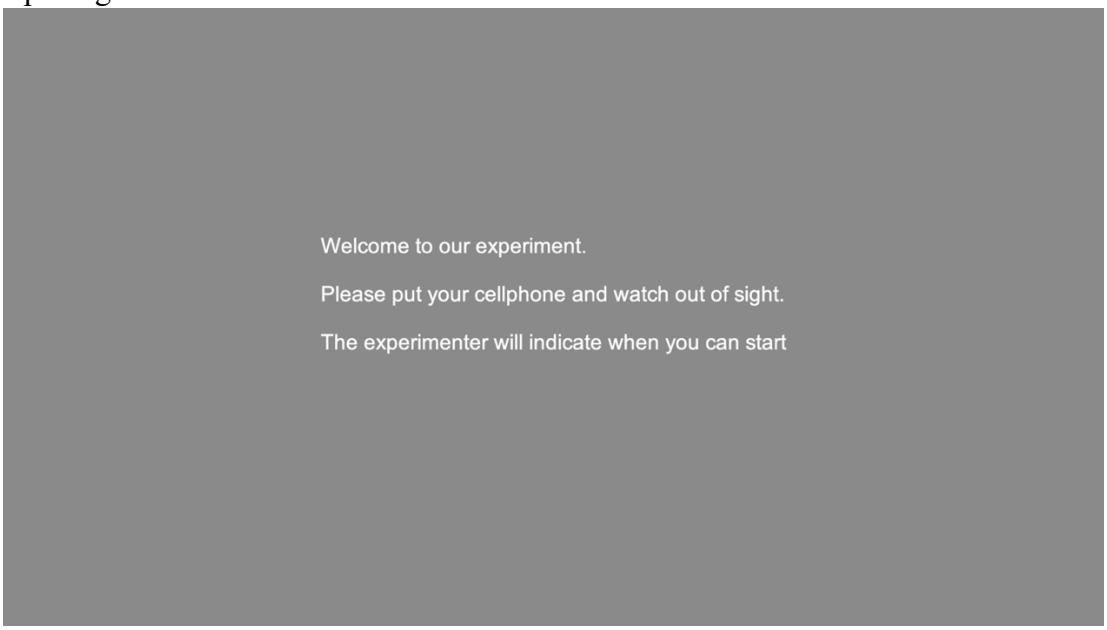

Figure 43: Opening Phase

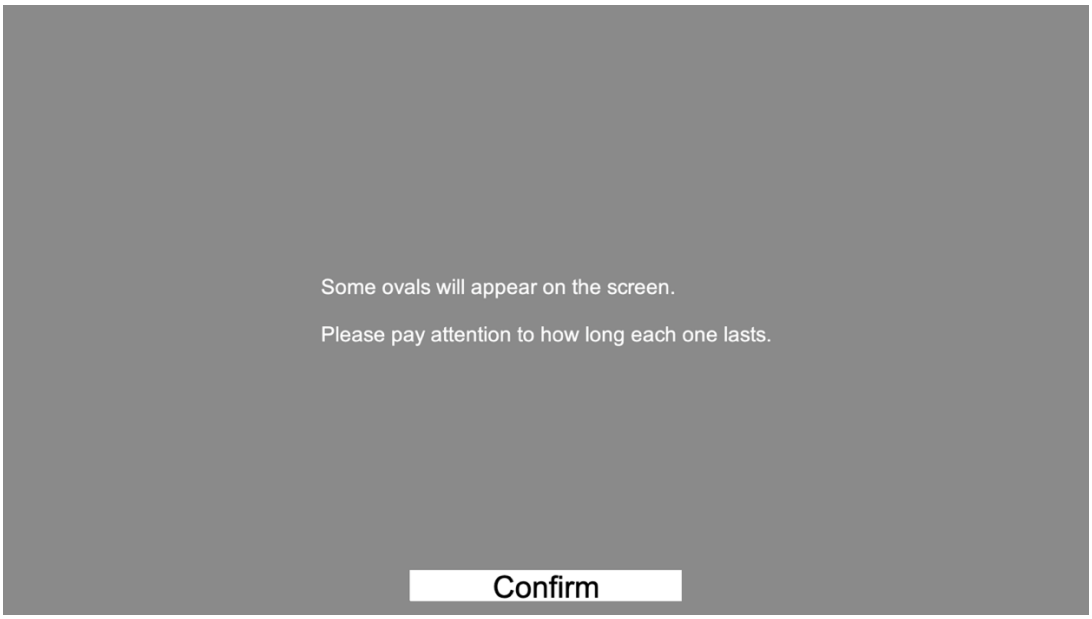

Figure 44: Sampling Phase

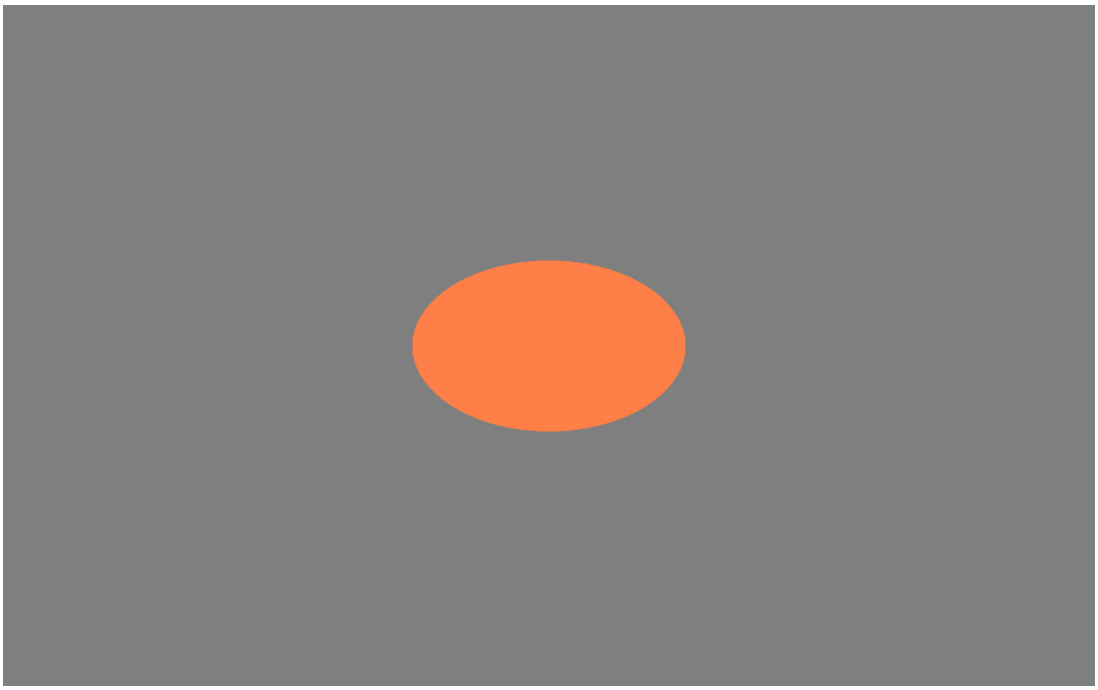

Figure 45: Sampling Phase – Experience

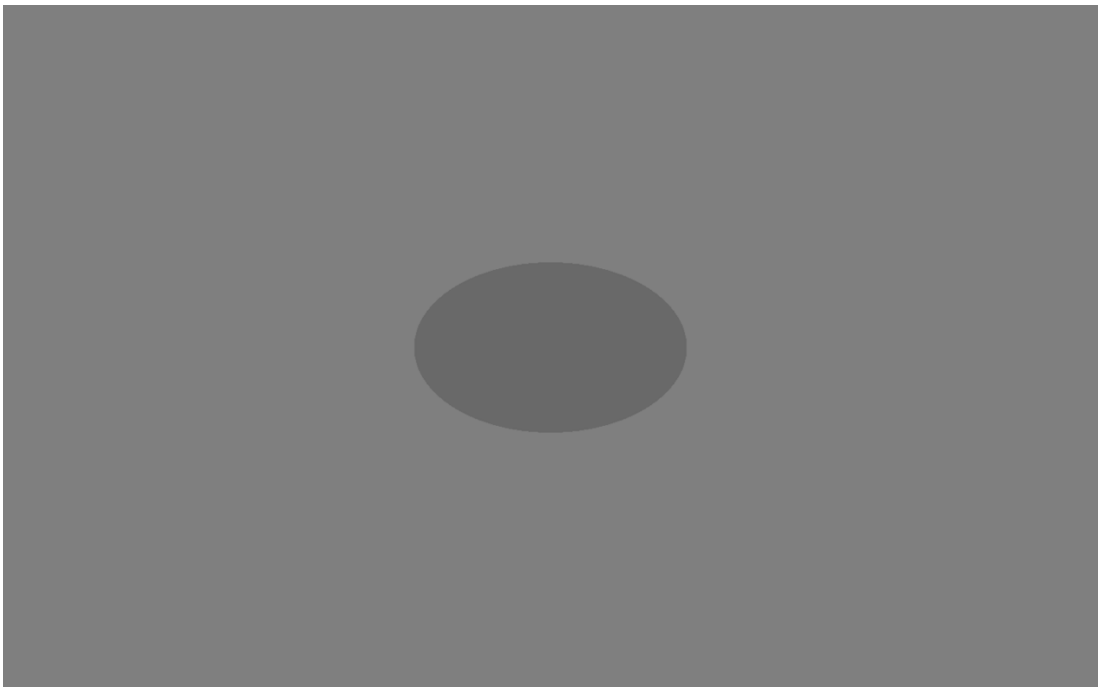

Figure 46: Sampling Phase – Description

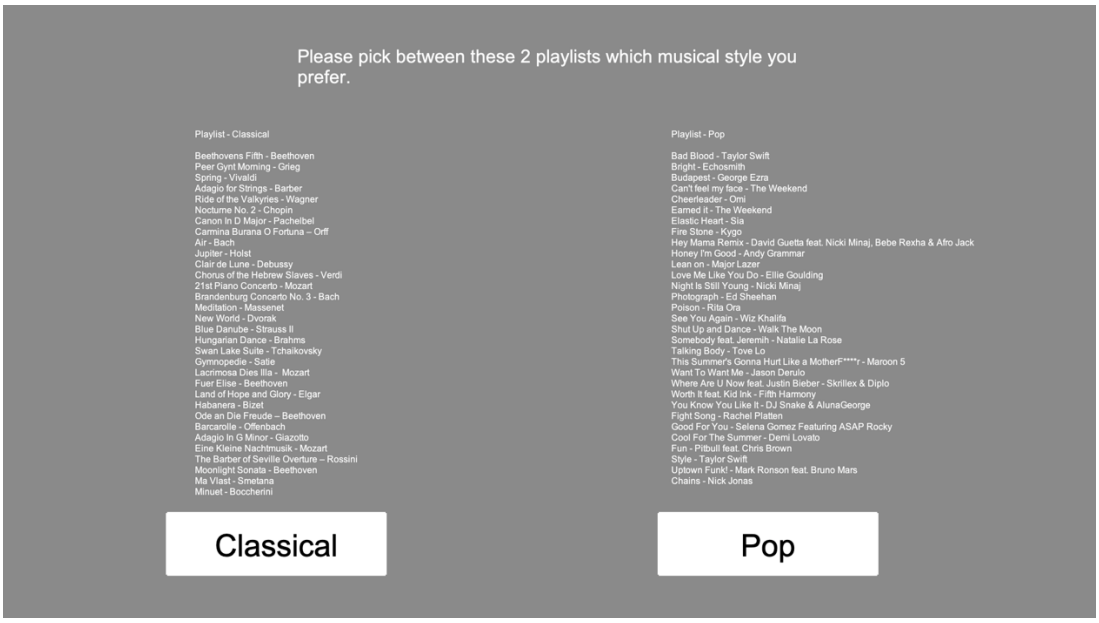

Figure 47: Reward Selection

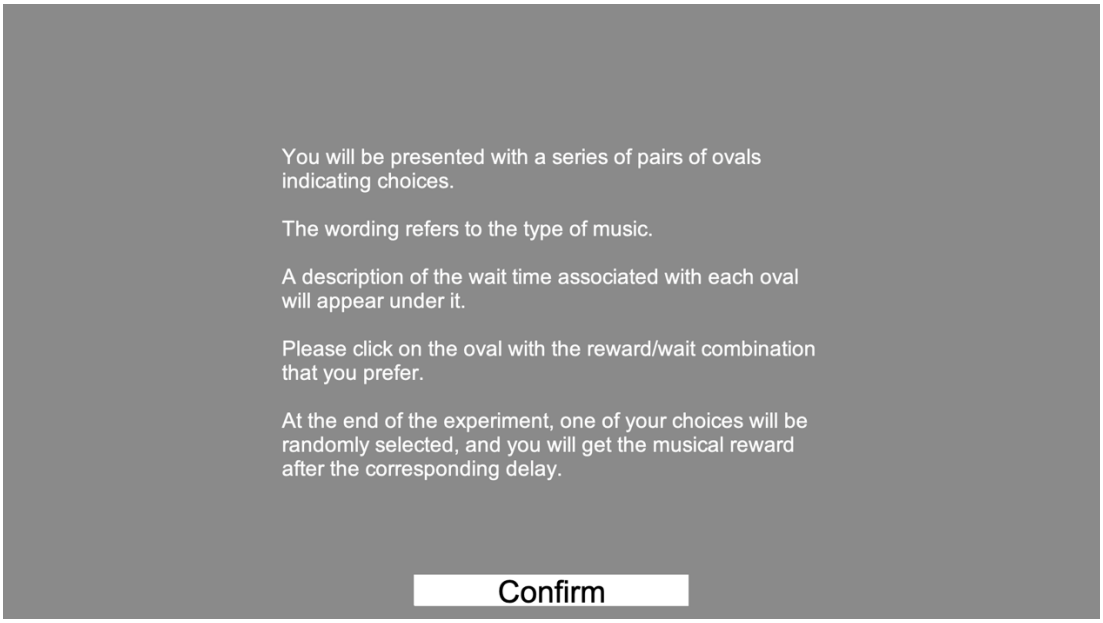

Figure 48: Decision Phase – Description

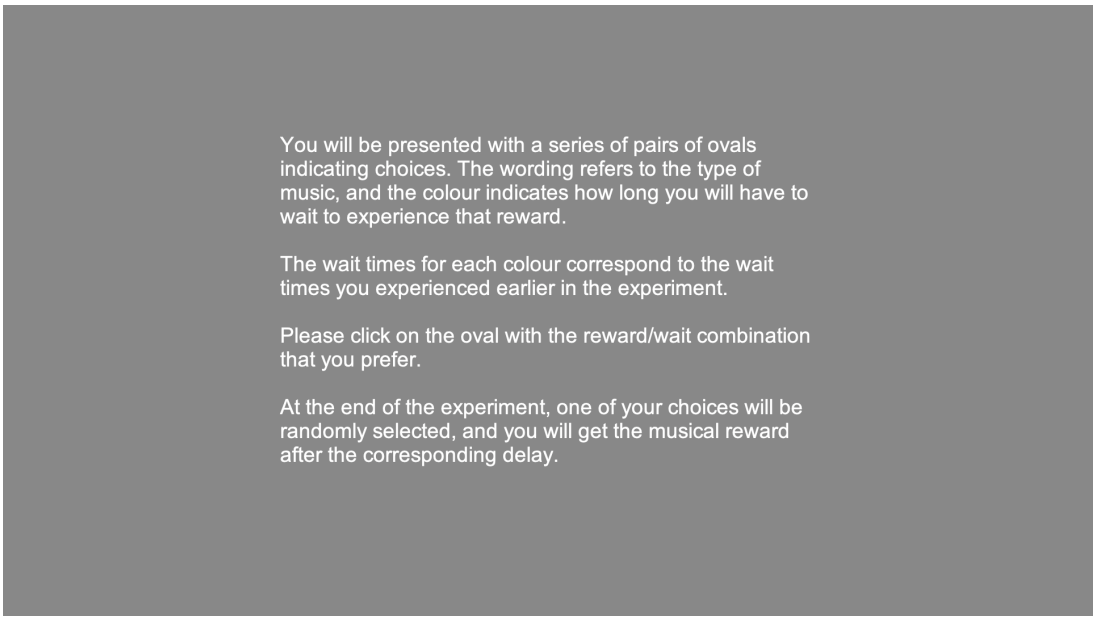

Figure 49: Decision Phase – Experience

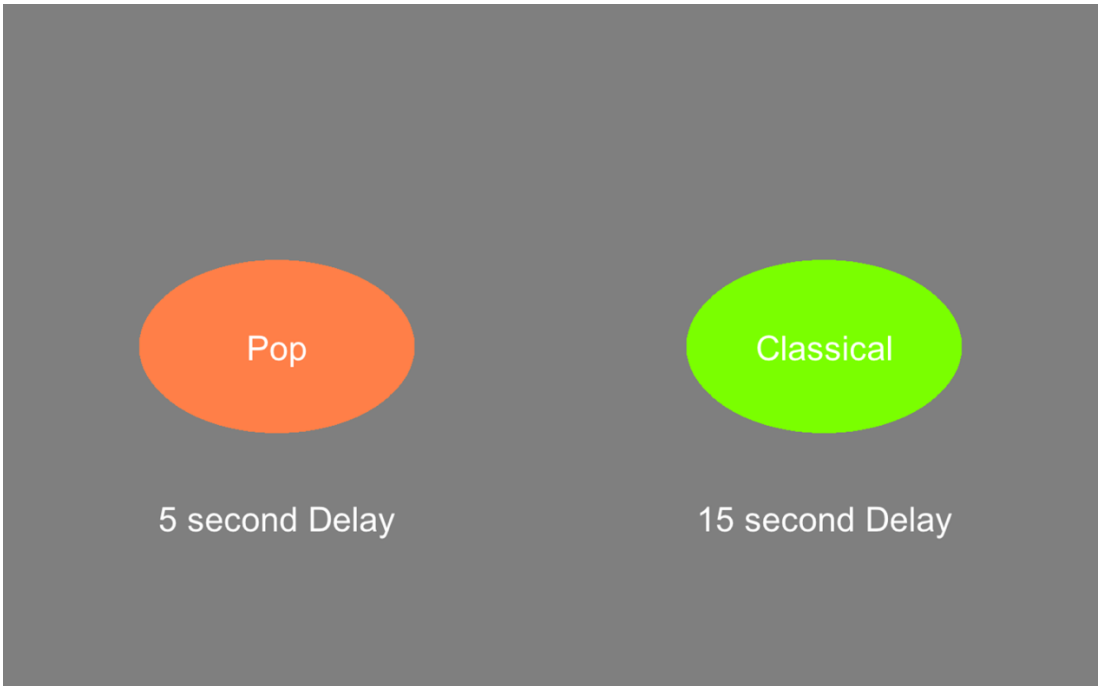

Figure 50: Decision Phase – Description

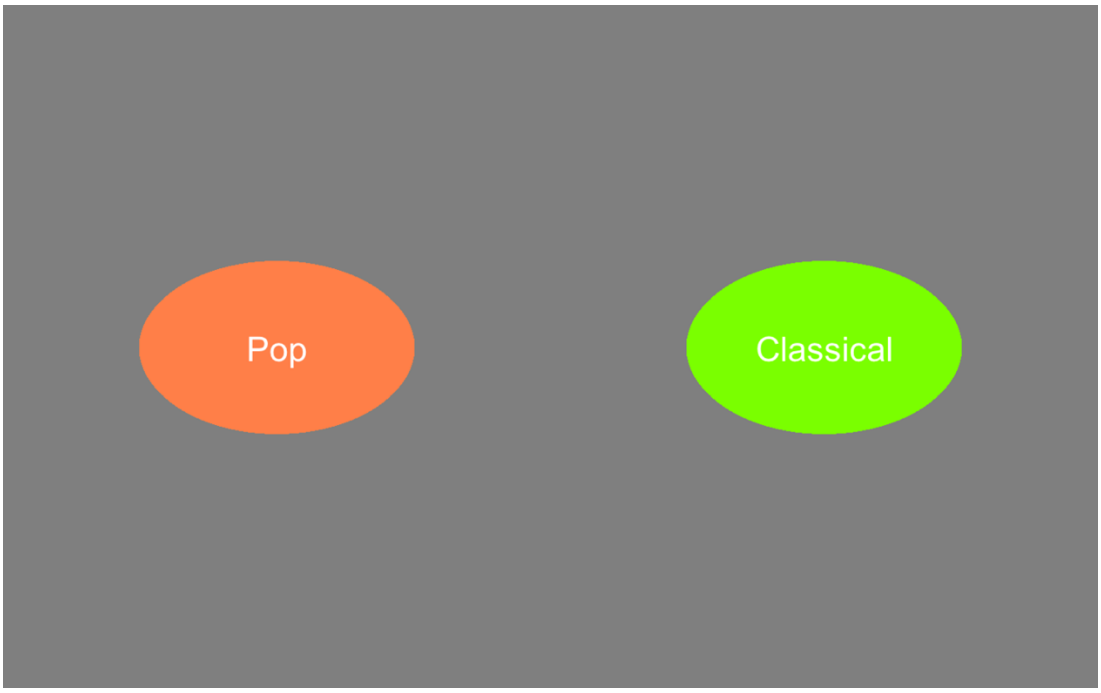

Figure 51: Decision Phase – Experience

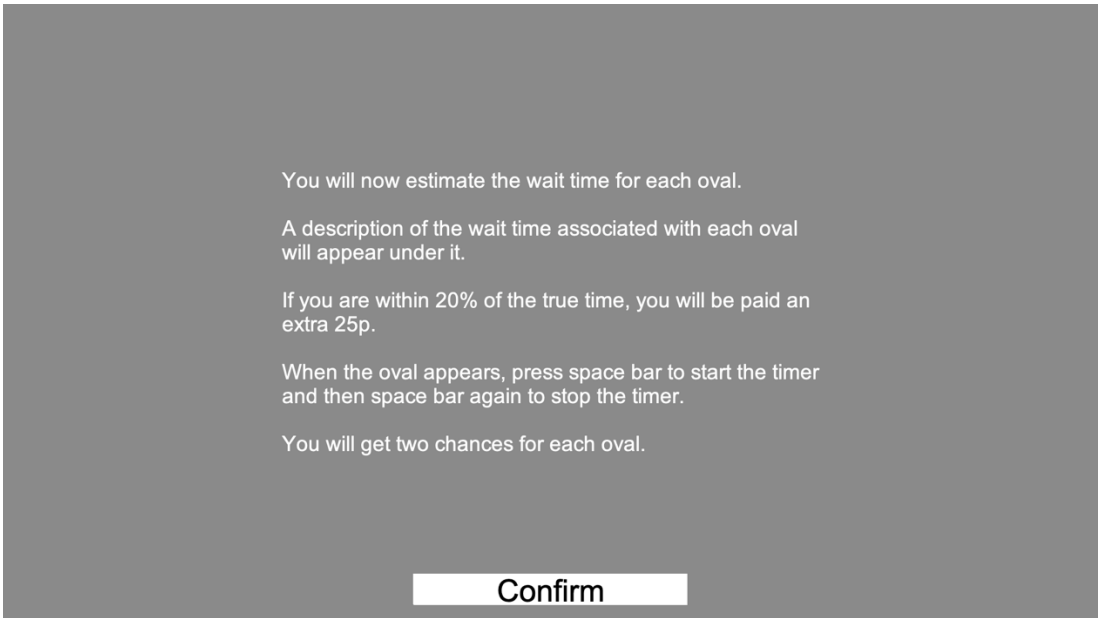

Figure 52: Time Estimation Phase – Description

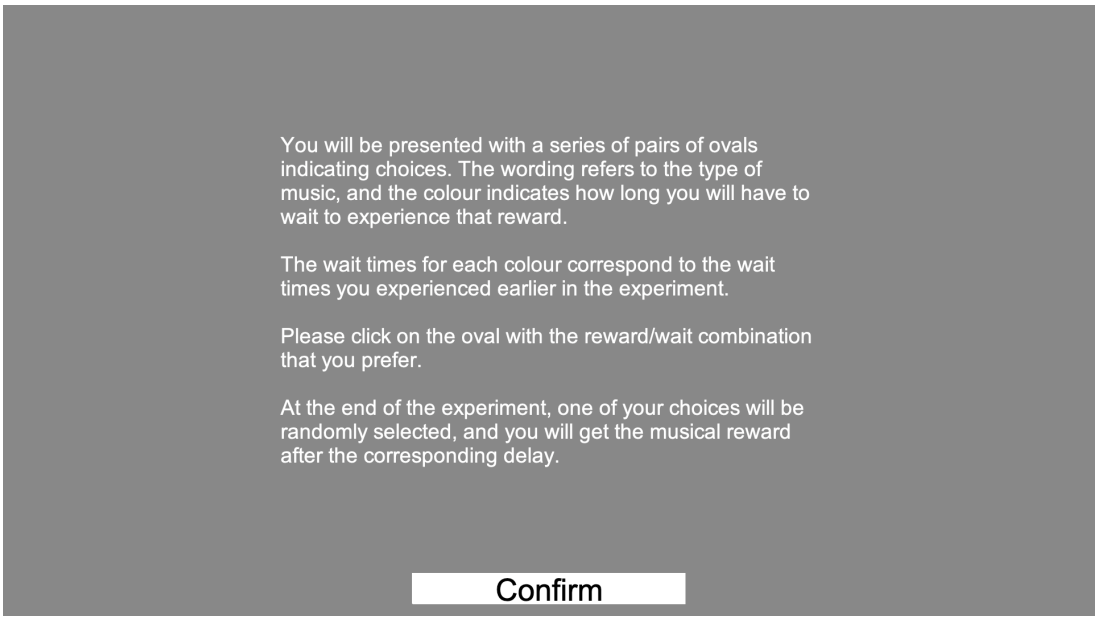

Figure 53: Time Estimation Phase – Experience

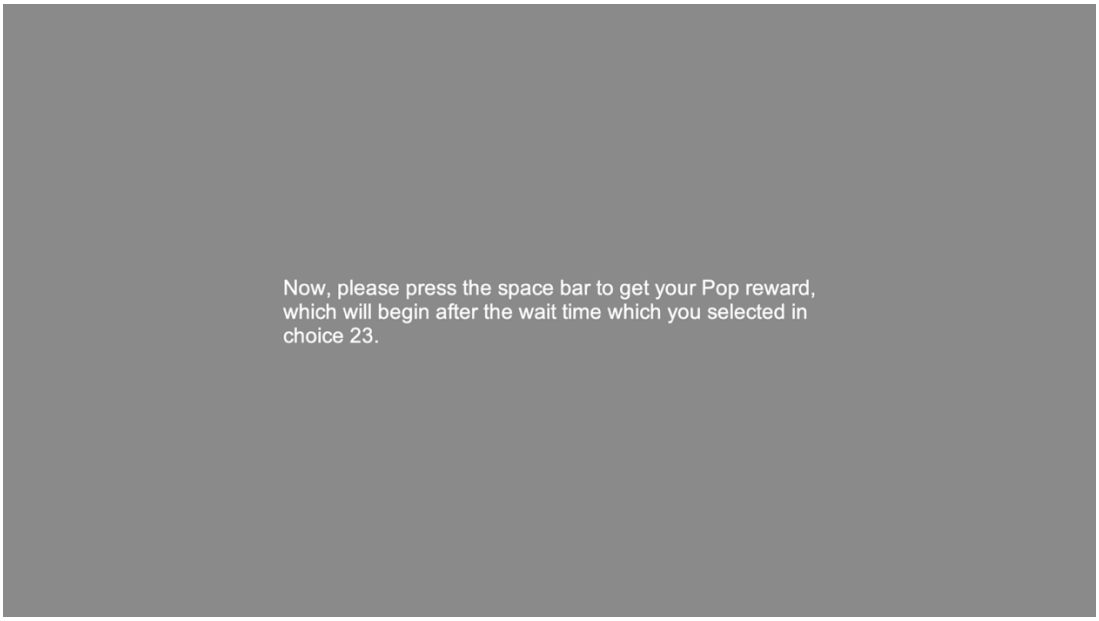

Figure 54: Reward Phase

Figure 55: Reward Phase

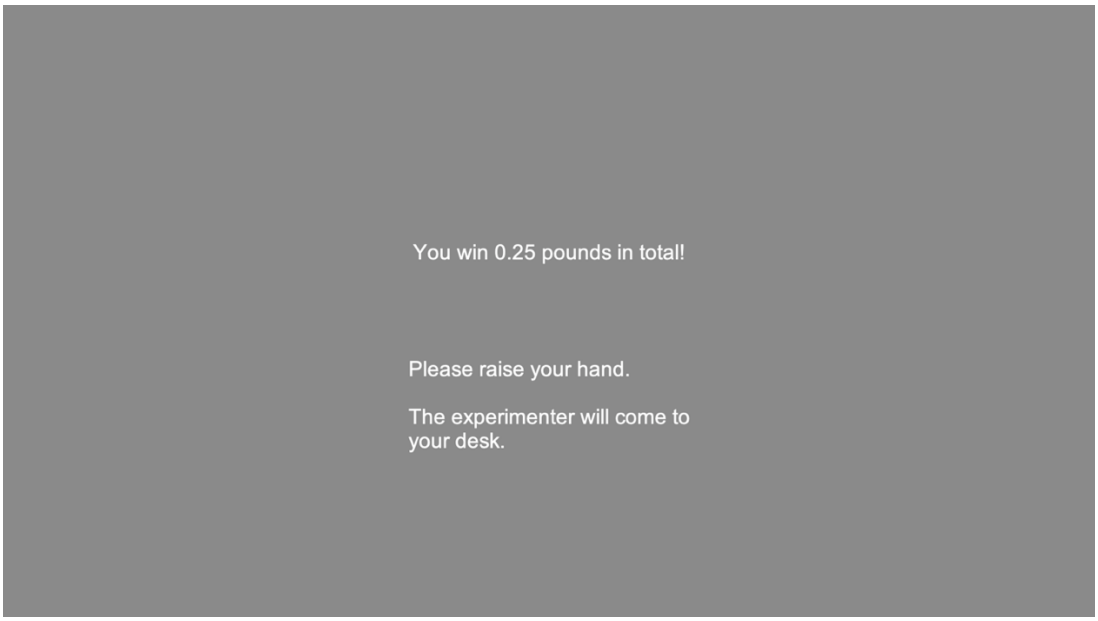

Figure 56: Reward Phase

Experiment 3

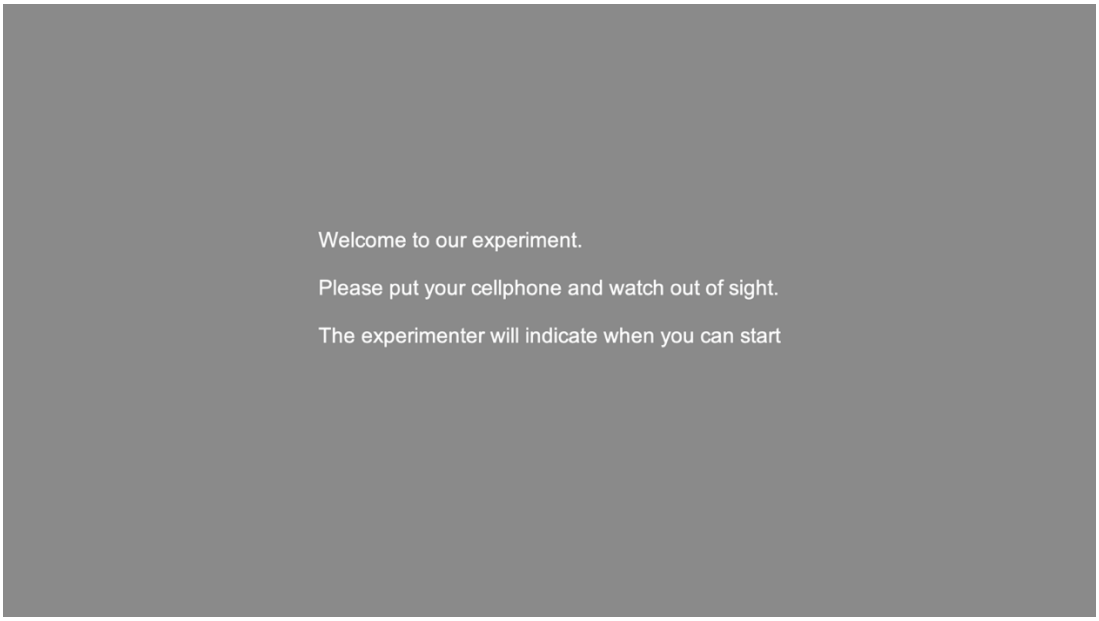

Figure 57: Opening Phase

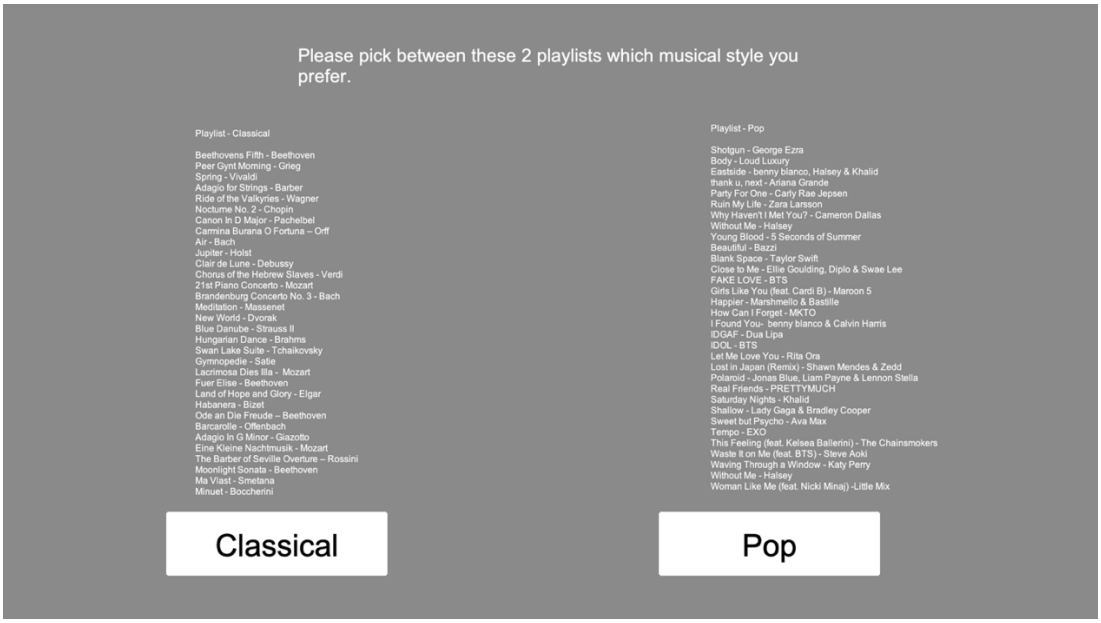

Figure 58: Reward Selection

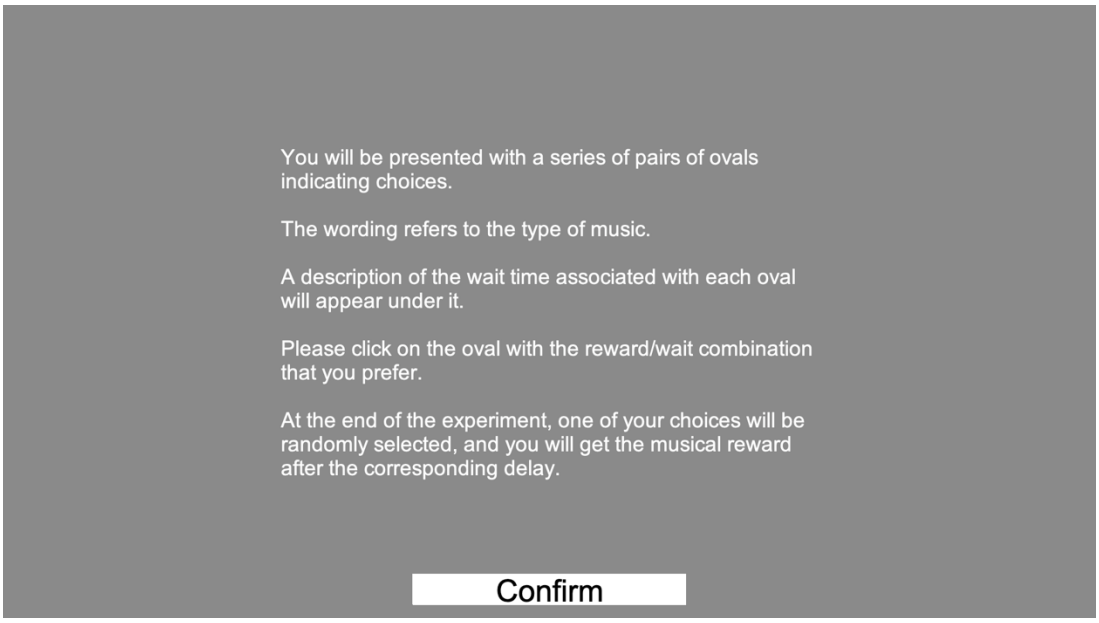

Figure 59: Decision Phase

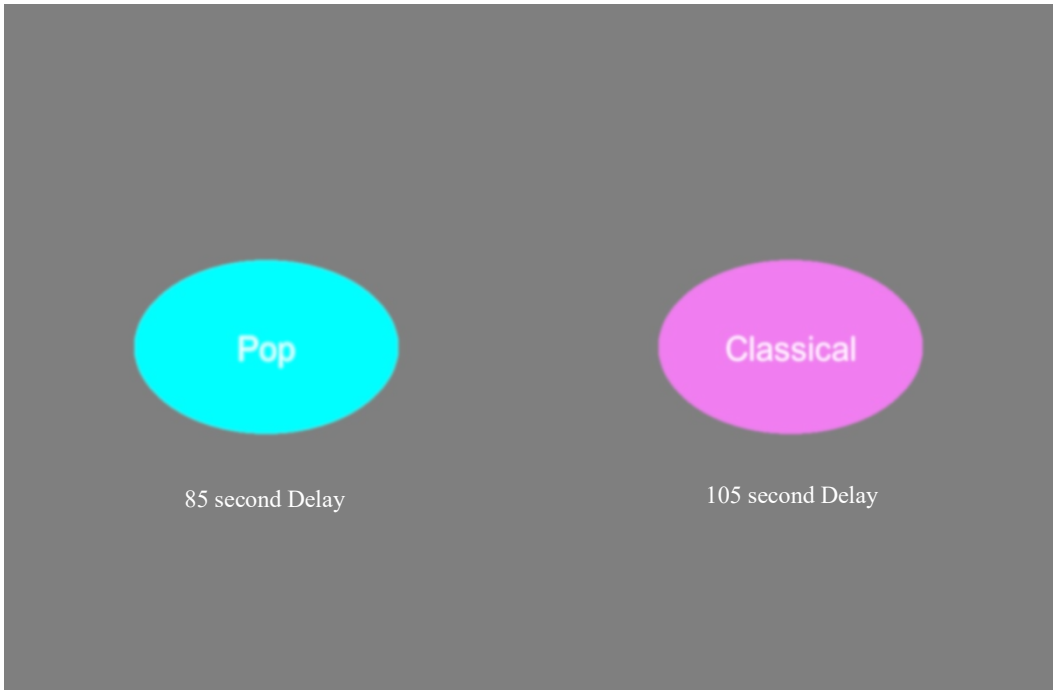

Figure 60: Decision Phase – Points

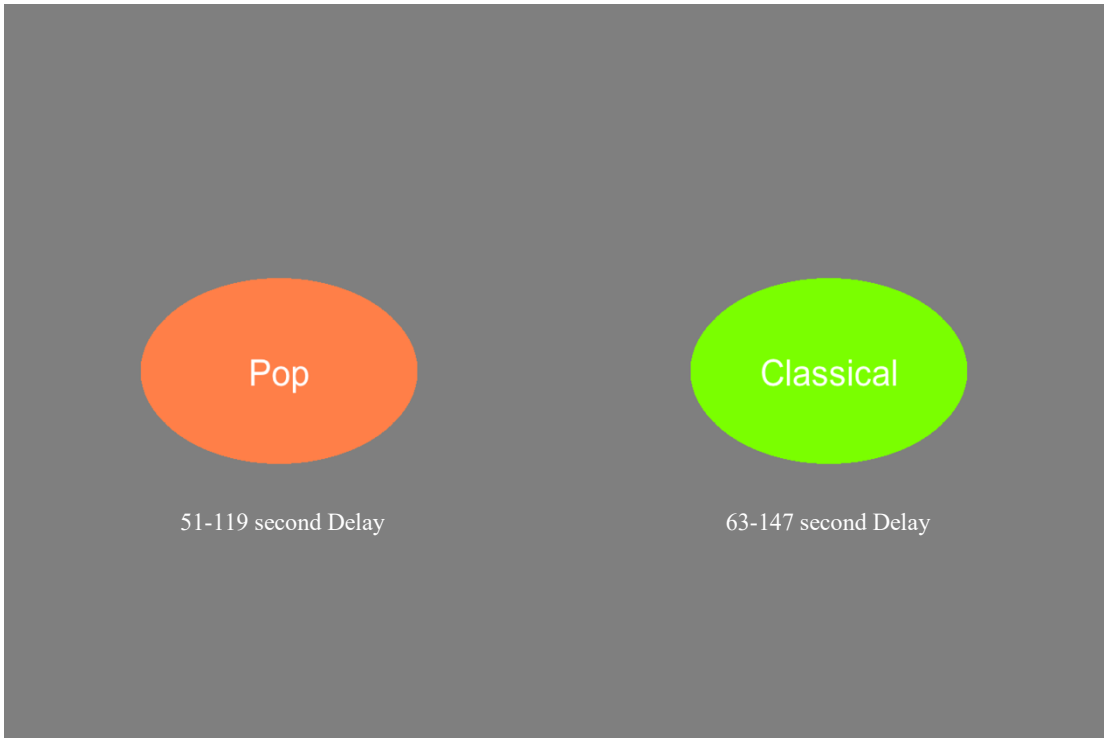

Figure 61: Decision Phase – Ranges

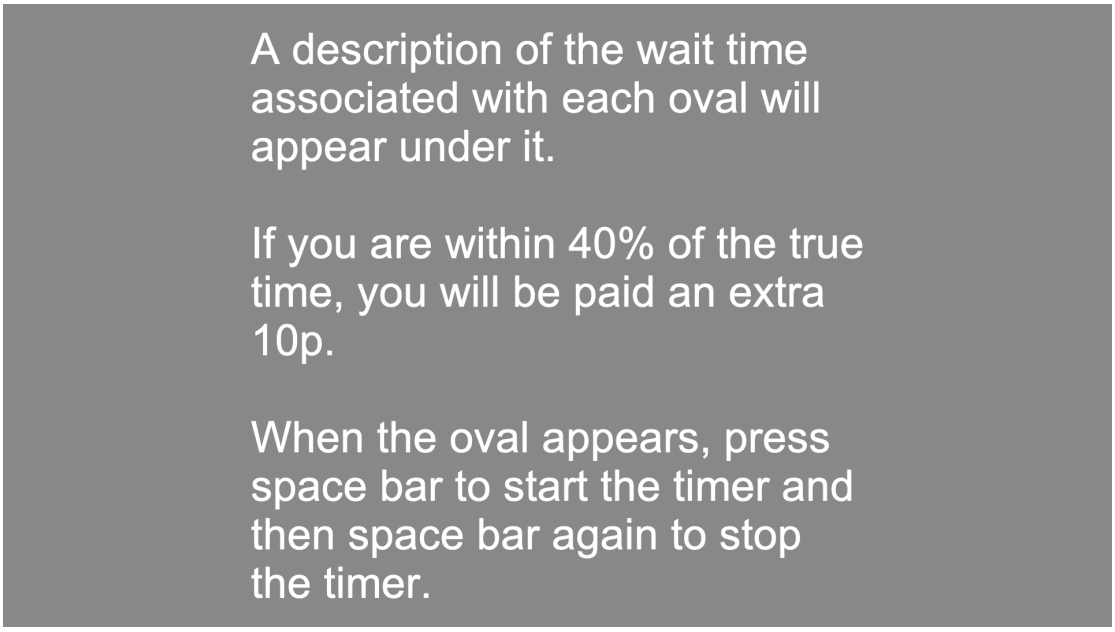

Figure 62: Time Estimation Phase – Points

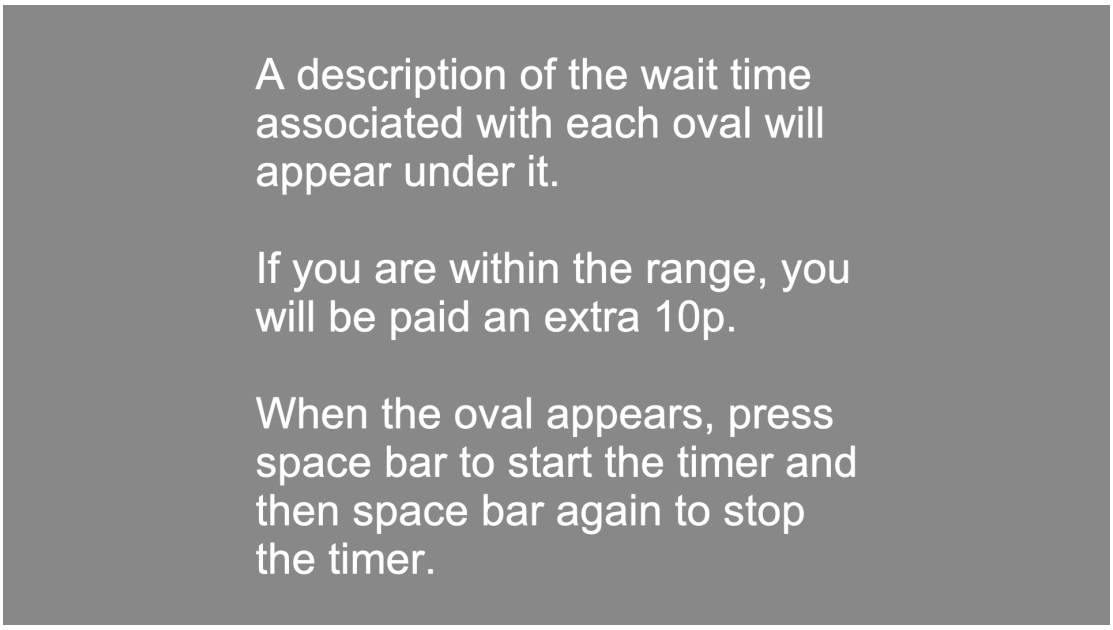

Figure 63: Time Estimation Phase – Ranges

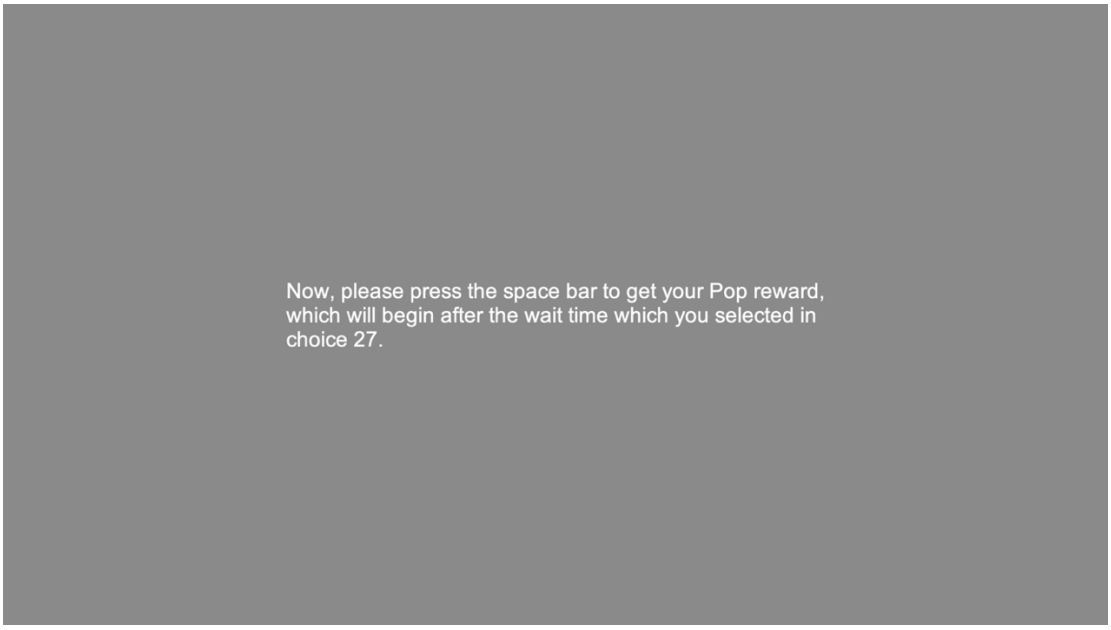

Figure 64: Reward Phase

Figure 65: Reward Phase

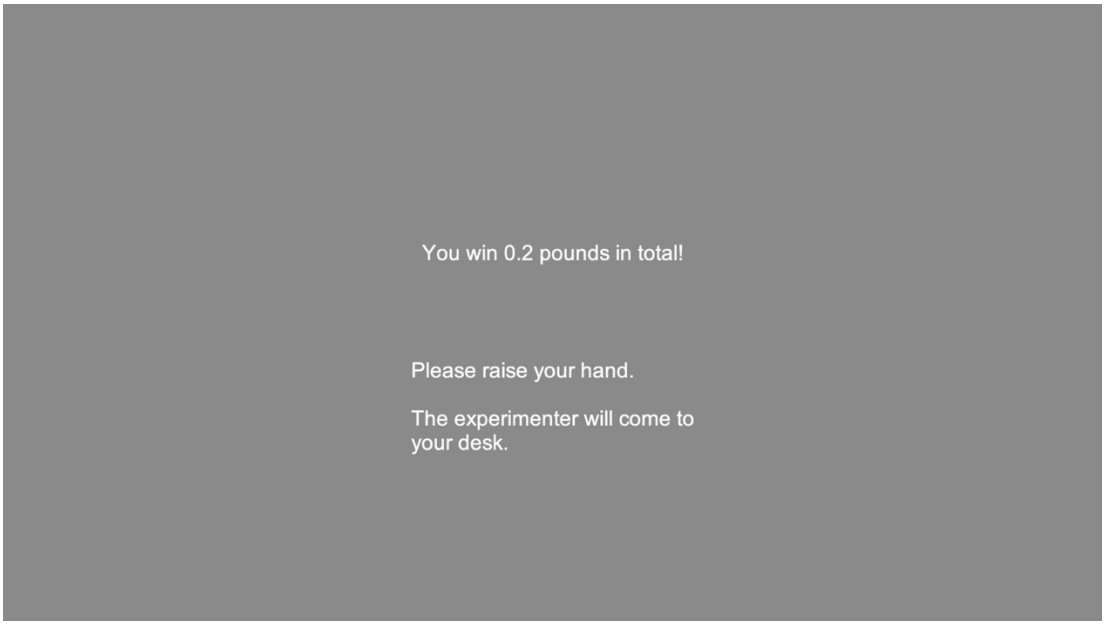

Figure 66: Reward Phase

Appendix 2

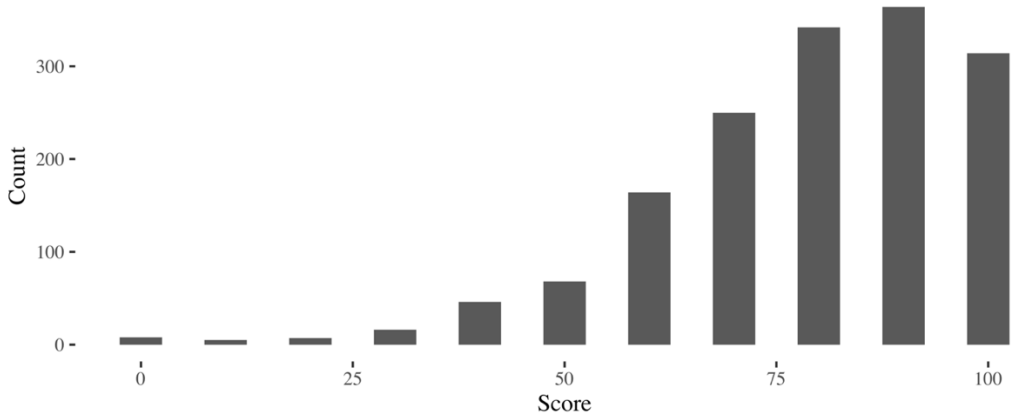

Figure 67: Histogram of Quiz Scores Collapsed Over Each Block

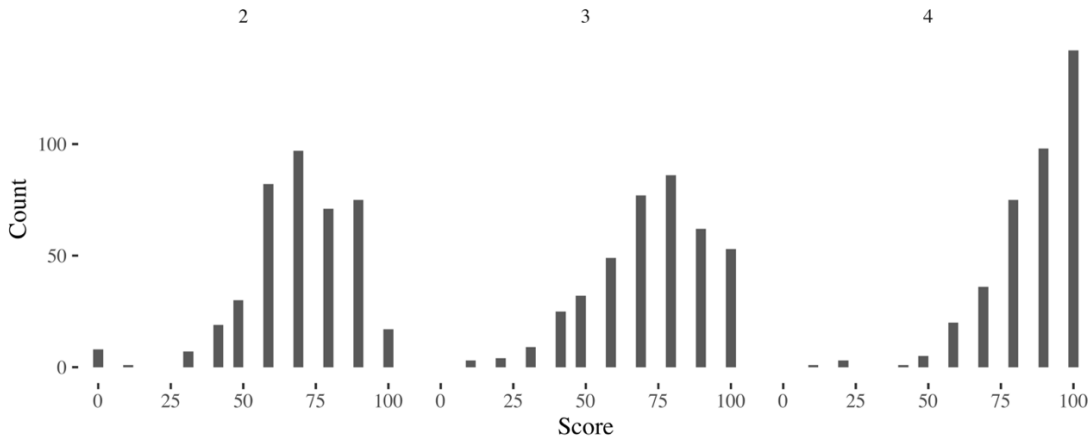

Figure 68: Histogram of Quiz Scores for Each Block

Bibliography

Ainslie, G. (1975) 'Specious reward: a behavioral theory of impulsiveness and impulse control.', *Psychological bulletin*, 82, pp. 463–496. doi: 10.1037/h0076860.

Andreoni, J. and Sprenger, C. (2012a) 'Estimating time preferences from convex budgets', *The American Economic Review*. American Economic Association, 102(7), pp. 3333–3356.

Andreoni, J. and Sprenger, C. (2012b) 'Risk Preferences Are Not Time Preferences', *American Economic Review*, 102(7), pp. 3357–3376. doi: 10.1257/aer.102.7.3357.

Ashby, N. J. S. and Gonzalez, C. (2017) 'The Influence of Time Estimation and Time-Saving Preferences on Learning to Make Temporally Dependent Decisions from Experience', *Journal of Behavioral Decision Making*. Wiley Online Library, 30(4), pp. 807–818.

Augenblick, N. (2017) 'Short-Term Time Discounting of Unpleasant Tasks'.

Augenblick, N., Niederle, M. and Sprenger, C. (2015) 'Working over Time: Dynamic Inconsistency in Real Effort Tasks', *The Quarterly Journal of Economics*, 130(3), pp. 1067–1115. doi: 10.1093/qje/qjv020.

Azmat, G. *et al.* (2019) 'What You Don't Know… Can't Hurt You? A Natural Field Experiment on Relative Performance Feedback in Higher Education', *Management Science*. INFORMS.

Azmat, G. and Iriberri, N. (2010) 'The importance of relative performance feedback information: Evidence from a natural experiment using high school students', *Journal of Public Economics*, 94(7), pp. 435–452. doi: https://doi.org/10.1016/j.jpubeco.2010.04.001.

Bandiera, O., Larcinese, V. and Rasul, I. (2015) 'Blissful ignorance? A natural experiment on the effect of feedback on students' performance', *Labour Economics*, 34, pp. 13–25. doi: https://doi.org/10.1016/j.labeco.2015.02.002.

Barron, G. and Erev, I. (2003) 'Small feedback-based decisions and their limited correspondence to description-based decisions', *Journal of Behavioral Decision Making*. Wiley Online Library, 16(3), pp. 215–233.

Becker, W. E. and Rosen, S. (1992) 'The learning effect of assessment and evaluation in high school', *Economics of Education Review*, 11(2), pp. 107–118. doi: https://doi.org/10.1016/0272-7757(92)90002-K.

Benhabib, J., Bisin, A. and Schotter, A. (2010) 'Present-bias, quasi-hyperbolic discounting, and fixed costs', *Games and Economic Behavior*, 69(2), pp. 205–223. doi: https://doi.org/10.1016/j.geb.2009.11.003.

Benzion, U., Rapoport, A. and Yagil, J. (1989) 'Discount Rates Inferred from Decisions: An Experimental Study', *Management Science*, pp. 270–284. doi: 10.1287/mnsc.35.3.270.

Bergman, P. and Chan, E. W. (2019) 'Leveraging Parents through Low-Cost Technology: The Impact of High-Frequency Information on Student Achievement'.

Bergman, P. and Rogers, T. (2017) 'The impact of defaults on technology adoption, and its underappreciation by policymakers'.

Bickel, W. K., Odum, A. L. and Madden, G. J. (1999) 'Impulsivity and cigarette smoking: delay discounting in current, never, and ex-smokers', *Psychopharmacology*, 146(4), pp. 447–454. doi: 10.1007/PL00005490.

Brandon, A. *et al.* (2017) *Do the effects of social nudges persist? Theory and evidence from 38 natural field experiments*. The Field Experiments Website.

Brocas, I., Carrillo, J. D. and Tarrasó, J. (2018) 'How long is a minute?', *Games and Economic Behavior*, 111, pp. 305–322. doi: https://doi.org/10.1016/j.geb.2018.06.007.

Brown, A. L., Chua, Z. E. and Camerer, C. F. (2009) 'Learning and Visceral Temptation in Dynamic Saving Experiments', *The Quarterly Journal of Economics*. JSTOR, 124(1), pp. 197–231. doi: 10.1162/qjec.2009.124.1.197.

Castleman, B. L. and Page, L. C. (2015) 'Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates?', *Journal of Economic Behavior & Organization*. Elsevier, 115, pp. 144–160.

Chande, R. *et al.* (2015) 'Curbing adult student attrition: Evidence from a field experiment'.

Chapman, G. B. (2000) 'Preferences for improving and declining sequences of health outcomes', *Journal of Behavioral Decision Making*. John Wiley & Sons, Ltd, 13(2), pp. 203–218. doi: 10.1002/(SICI)1099-0771(200004/06)13:2<203::AID-BDM317>3.0.CO;2-S.

Chapman, G. B. and Elstein, A. S. (1995) 'Valuing the Future: Temporal Discounting of Health and Money', *Medical Decision Making*. SAGE Publications Inc STM, 15(4), pp. 373–386. doi:

10.1177/0272989X9501500408.

Chapman, G. B. and Weber, B. J. (2006) 'Decision biases in intertemporal choice and choice under uncertainty: testing a common account', *Memory & Cognition*. Springer, 34(3), pp. 589–602.

Chaturapruek, S. *et al.* (2018) 'How a data-driven course planning tool affects college students' GPA: Evidence from two field experiments', *Proceedings of the Fifth ACM Conference on Learning at Scale*, p. in press. doi: 10.1145/3231644.3231668.

Cohen, J. *et al.* (2016) *Measuring Time Preferences*. Cambridge, MA. doi: 10.3386/w22455.

Crockett, M. J. *et al.* (2013) 'Restricting temptations: neural mechanisms of precommitment', *Neuron*. Elsevier, 79(2), pp. 391–401.

Dai, J. *et al.* (2019) 'What the Future Holds and When: A Description–Experience Gap in Intertemporal Choice', *Psychological science*. SAGE Publications Sage CA: Los Angeles, CA, p. 0956797619858969.

Davis, D. *et al.* (2017) 'Follow the Successful Crowd: Raising MOOC Completion Rates Through Social Comparison at Scale', in *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. New York, NY, USA: ACM (LAK '17), pp. 454–463. doi: 10.1145/3027385.3027411.

Doss, C. *et al.* (2017) 'Supporting Parenting through Differentiated and Personalized Text-Messaging: Testing Effects on Learning during Kindergarten. CEPA Working Paper No. 16-18.', *Stanford Center for Education Policy Analysis*. ERIC.

Dubey, P. and Geanakoplos, J. (2010) 'Grading exams: 100,99,98,… or A,B,C?', *Games and Economic Behavior*, 69(1), pp. 72–94. doi: https://doi.org/10.1016/j.geb.2010.02.001.

Dutt, V., Arló-Costa, H., Helzner, J., & Gonzalez, C. (2014). The Description–Experience Gap in Risky and Ambiguous Gambles. *Journal of Behavioral Decision Making*, *27*(4), 316–327. https://doi.org/10.1002/bdm.1808

Dweck, C. S. (2013) *Self-theories: Their role in motivation, personality, and development*. Psychology press.

Elsner, B. and Isphording, I. E. (2016) 'A Big Fish in a Small Pond: Ability Rank and Human Capital Investment', *Journal of Labor Economics*. The University of Chicago Press, 35(3), pp. 787–828. doi: 10.1086/690714.

Ert, E. and Lejarraga, T. (2018) 'The Effect of Experience on Context-dependent

Decisions', *Journal of Behavioral Decision Making*. John Wiley & Sons, Ltd, 31(4), pp. 535–546. doi: 10.1002/bdm.2064.

Estle, S. J. *et al.* (2006) 'Differential effects of amount on temporal and probability discounting of gains and losses', *Memory & Cognition*, 34(4), pp. 914–928. doi: 10.3758/BF03193437.

Estle, S. J. *et al.* (2007) 'Discounting of Monetary and Directly Consumable Rewards', *Psychological Science*. SAGE Publications Inc, 18(1), pp. 58–63. doi: 10.1111/j.1467-9280.2007.01849.x.

Fisher, I. (1930) *The Theory of interest*. NY: Macmillan.

Fox, C. R. and Hadar, L. (2006) 'Decisions from experience= sampling error+ prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004)', *Judgment and Decision Making*. Society for Judgment and Decision Making, 1(2), pp. 159–161.

Frederick, S. *et al.* (2002) 'Time Discounting and Preference : A Critical Review', *Journal of Economic Literature*, 40, pp. 351–401. doi: 10.1126/science.151.3712.867-a.

Furlong, E. E. and Opfer, J. E. (2009) 'Cognitive constraints on how economic rewards affect cooperation', *Psychological Science*. SAGE Publications Sage CA: Los Angeles, CA, 20(1), pp. 11–16.

Gonzalez, C. (2013). Chapter 5 - The boundaries of instance-based learning theory for explaining decisions from experience. In V. S. C. Pammi & N. Srinivasan (Eds.), *Progress in Brain Research* (Vol. 202, pp. 73–98). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-444-62604-2.00005-8

Harman, J. L., & Gonzalez, C. (2015). *Allais from Experience: Choice Consistency, Rare Events, and Common Consequences in Repeated Decisions*. https://doi.org/10.1002/bdm.1855

Gallistel, C. R. and Gibbon, J. (2000) 'Time, rate, and conditioning.', *Psychological review*. American Psychological Association, 107(2), p. 289.

Gill, D. *et al.* (2018) 'First-Place Loving and Last-Place Loathing: How Rank in the Distribution of Performance Affects Effort Provision', *Management Science*, p. mnsc.2017.2907. doi: 10.1287/mnsc.2017.2907.

Green, L., Fristoe, N. and Myerson, J. (1994) 'Temporal discounting and preference reversals in choice between delayed outcomes', *Psychonomic Bulletin & Review*, 1(3), pp. 383–389. doi: 10.3758/BF03213979.

Green, L. and Myerson, J. (2004) 'A Discounting Framework for Choice With

Delayed and Probabilistic Rewards.', *Psychological Bulletin*.

Hadar, L., Danziger, S. and Hertwig, R. (2018) 'The Attraction Effect in Experience-based Decisions', *Journal of Behavioral Decision Making*. Wiley Online Library.

Han, R. and Takahashi, T. (2012) 'Psychophysics of time perception and valuation in temporal discounting of gain and loss', *Physica A: Statistical Mechanics and its Applications*, 391(24), pp. 6568–6576. doi: https://doi.org/10.1016/j.physa.2012.07.012.

Heilbronner, S. R. and Hayden, B. Y. (2015) 'The description-experience gap in risky choice in nonhuman primates', *Psychonomic bulletin & review*. Springer, pp. 1–8.

Hertwig, R. *et al.* (2004) 'Decisions from experience and the effect of rare events in risky choice', *Psychological Science*. SAGE Publications, 15(8), pp. 534–539.

Hertwig, R. and Erev, I. (2009) 'The description–experience gap in risky choice', *Trends in cognitive sciences*. Elsevier, 13(12), pp. 517–523.

Holt, D. D. *et al.* (2008) 'Preference reversals with losses', *Psychonomic Bulletin & Review*, 15(1), pp. 89–95. doi: 10.3758/PBR.15.1.89.

Hyman, I. E. *et al.* (2009) 'Did you see the unicycling clown? Inattentional blindness while walking and talking on a cell phone', *Applied Cognitive Psychology*, 24(5), pp. 597–607. doi: 10.1002/acp.1638.

Ikink, I. *et al.* (2019) 'Time ambiguity during intertemporal decision-making is aversive, impacting choice and neural value coding', *NeuroImage*. doi: 10.1016/j.neuroimage.2018.10.008.

Imai, T., Rutter, T. and Camerer, C. (2019) 'Meta-Analysis of Present-Bias Estimation using Convex Time Budgets'. MetaArXiv.

Jersild, A. T. (1927) 'Mental set and shift.', *Archives of Psychology*, 14, 89, p. 81.

Johnson, M. W. (2012) 'An efficient operant choice procedure for assessing delay discounting in humans: Initial validation in cocaine-dependent and control individuals.', *Experimental and clinical psychopharmacology*. American Psychological Association, 20(3), p. 191.

Kahneman, D. and Tversky, A. (1979) 'Prospect Theory: An Analysis of Decision under Risk', *Econometrica*, 47, pp. 263–291. doi: 10.2307/1914185.

Keren, G. and Roelofsma, P. (1995) 'Immediacy and Certainty in Intertemporal Choice', *Organizational Behavior and Human Decision Processes*, 63(3),

pp. 287–297. doi: https://doi.org/10.1006/obhd.1995.1080.

Kidd, C., Palmeri, H. and Aslin, R. N. (2013) 'Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability', *Cognition*. Elsevier, 126(1), pp. 109–114.

Kirby, K. N. and Herrnstein, R. J. (1995) 'Preference Reversals Due to Myopic Discounting of Delayed Reward', *Psychological Science*. SAGE Publications Inc, 6(2), pp. 83–89. doi: 10.1111/j.1467-9280.1995.tb00311.x.

Knight, F. H. (1921) *Risk, Uncertainty, and Profit*. New York: Houghton Mifflin.

Laibson, D. (1997) 'Golden eggs and hyperbolic discounting', *The Quarterly Journal of Economics*, 112, pp. 443–478. doi: 10.1162/003355397555253.

Leland, J. W. (1994). Generalized similarity judgments: An alternative explanation for choice anomalies. *Journal of Risk and Uncertainty*, *9*(2), 151–172. https://doi.org/10.1007/BF01064183

Loewenstein, G. (1987). Anticipation and the Valuation of Delayed Consumption. *The Economic Journal*, *97*(387), 666–684. https://doi.org/10.2307/2232929

Loewenstein, G. F. and Prelec, D. (1993) 'Preferences for sequences of outcomes.', *Psychological Review*. US: American Psychological Association, 100(1), pp. 91–108. doi: 10.1037/0033-295X.100.1.91.

Loewenstein, G. and Prelec, D. (1992) 'Anomalies in Intertemporal Choice: Evidence and an Interpretation*', *The Quarterly Journal of Economics*, 107(2), pp. 573–597. doi: 10.2307/2118482.

Loewenstein, G. and Sicherman, N. (1991) 'Do Workers Prefer Increasing Wage Profiles?', *Journal of Labor Economics*. The University of Chicago Press, 9(1), pp. 67–84. doi: 10.1086/298259.

Luckman, A., Donkin, C. and Newell, B. R. (2018) 'Can a single model account for both risky choices and inter-temporal choices? Testing the assumptions underlying models of risky inter-temporal choice', *Psychonomic Bulletin & Review*, 25(2), pp. 785–792. doi: 10.3758/s13423-017-1330-8.

Ludvig, E. A. *et al.* (2018) 'Living near the edge: How extreme outcomes and their neighbors drive risky choice.', *Journal of Experimental Psychology: General*. American Psychological Association.

Ludvig, E. A. and Spetch, M. L. (2011) 'Of black swans and tossed coins: Is the description-experience gap in risky choice limited to rare events?', *PLoS ONE*, 6(6). doi: 10.1371/journal.pone.0020262.

Lukinova, E. *et al.* (2019) 'Time preferences are reliable across time-horizons and verbal versus experiential tasks', *eLife*. eLife Sciences Publications Limited, 8, p. e39656.

Magen, E., Dweck, C. S. and Gross, J. J. (2008) 'The hidden-zero effect: representing a single choice as an extended sequence reduces impulsive choice', *Psychological science*, 19(7), pp. 648–649. doi: 10.1111/j.1467-9280.2008.02137.x.

Main, J. B. and Ost, B. (2014) 'The Impact of Letter Grades on Student Effort, Course Selection, and Major Choice: A Regression-Discontinuity Analysis', *The Journal of Economic Education*. Routledge, 45(1), pp. 1–10. doi: 10.1080/00220485.2014.859953.

Maltby, J. *et al.* (2016) 'Social ranking effects on tooth-brushing behaviour', *British Journal of Health Psychology*, 21, pp. 374–388. doi: 10.1111/bjhp.12173.

Marzilli Ericson, K. M. *et al.* (2015) 'Money Earlier or Later? Simple Heuristics Explain Intertemporal Choices Better Than Delay Discounting Does', *Psychological Science*. SAGE Publications, 26(6), pp. 826–833. doi: 10.1177/0956797615572232.

Matsakis, L. (2019) *At an Outback Steakhouse Franchise, Surveillance Blooms | WIRED*. Available at: https://www.wired.com/story/outback-steakhouse-presto-vision-surveillance/?mbid=social_twitter&utm_brand=wired&utm_campaign=wired&utm_medium=social&utm_social-type=owned&utm_source=twitter (Accessed: 11 November 2019).

Mazur, J. E. (1987) 'An adjusting procedure for studying delayed reinforcement', *Commons, ML.; Mazur, JE.; Nevin, JA*, pp. 55–73.

McClure, S. M. *et al.* (2004) 'Separate neural systems value immediate and delayed monetary rewards.', *Science (New York, N.Y.)*, 306, pp. 503–507. doi: 10.1126/science.1100907.

McClure, S. M. *et al.* (2007) 'Time discounting for primary rewards', *The Journal of Neuroscience*. Soc Neuroscience, 27(21), pp. 5796–5804.

McGuire, J. T. and Kable, J. W. (2012) 'Decision makers calibrate behavioral persistence on the basis of time-interval experience', *Cognition*. Elsevier, 124(2), pp. 216–226.

McGuire, J. T. and Kable, J. W. (2013) 'Rational temporal predictions can underlie apparent failures to delay gratification.', *Psychological review*. American Psychological Association, 120(2), p. 395.

McGuire, J. T. and Kable, J. W. (2015) 'Medial prefrontal cortical activity reflects

dynamic re-evaluation during voluntary persistence', *Nature neuroscience*. Nature Publishing Group, 18(5), pp. 760–766.

Michaelson, L. *et al.* (2013) 'Delaying gratification depends on social trust ', *Frontiers in Psychology* , p. 355. Available at: https://www.frontiersin.org/article/10.3389/fpsyg.2013.00355.

Mischel, W. and Ebbesen, E. B. (1970) 'Attention in delay of gratification.', *Journal of Personality and Social Psychology*. American Psychological Association, 16(2), p. 329.

Mischel, W., Shoda, Y. and Rodriguez, M. I. (1989) 'Delay of gratification in children', *Science*. American Association for the Advancement of Science, 244(4907), pp. 933–938.

Murphy, R. and Weinhardt, F. (2018) 'Top of the Class: The Importance of Ordinal Rank', *National Bureau of Economic Research Working Paper Series*, No. 24958. doi: 10.3386/w24958.

Natriello, G. and McDill, E. L. (1986) 'Performance Standards, Student Effort on Homework, and Academic Achievement', *Sociology of Education*. [Sage Publications, Inc., American Sociological Association], 59(1), pp. 18–31. doi: 10.2307/2112483.

NHTSA (2017) *Distracted Driving | NHTSA*, *Nhtsa.Gov*. Available at: https://www.nhtsa.gov/risky-driving/distracted-driving (Accessed: 7 November 2019).

O'Donoghue, T. and Rabin, M. (1999a) 'Doing It Now or Later', *American Economic Review*, 89(1), pp. 103–124. doi: 10.1257/aer.89.1.103.

O'Donoghue, T. and Rabin, M. (1999b) 'Incentives for Procrastinators*', *The Quarterly Journal of Economics*, 114(3), pp. 769–816. doi: 10.1162/003355399556142.

O'Donoghue, T. and Rabin, M. (2015) 'Present Bias: Lessons Learned and to Be Learned', *American Economic Review*. American Economic Association, 105(5), pp. 273–279.

Odum, A. L., Baumann, A. A. L. and Rimington, D. D. (2006) 'Discounting of delayed hypothetical money and food: Effects of amount', *Behavioural Processes*, 73(3), pp. 278–284. doi: https://doi.org/10.1016/j.beproc.2006.06.008.

Odum, A. L. and Rainaud, C. P. (2003) 'Discounting of delayed hypothetical money, alcohol, and food', *Behavioural Processes*, 64(3), pp. 305–313. doi: https://doi.org/10.1016/S0376-6357(03)00145-1.

Paredes, V. (2016) ‘Grading System and Student Effort’, *Education Finance and Policy*. MIT Press, 12(1), pp. 107–128. doi: 10.1162/EDFP_a_00195.

Phelps, E. S. and Pollak, R. A. (1968) ‘On Second-Best National Saving and Game-Equilibrium Growth’, *The Review of Economic Studies*. [Oxford University Press, Review of Economic Studies, Ltd.], 35(2), pp. 185–199. doi: 10.2307/2296547.

Plonsky, O., & Erev, I. (2017). Learning in settings with partial feedback and the wavy recency effect of rare events. *Cognitive Psychology*, *93*, 18–43. https://doi.org/https://doi.org/10.1016/j.cogpsych.2017.01.002

Plonsky, O., Teodorescu, K., & Erev, I. (2015). Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychological Review*, *122*(4), 621–647. https://doi.org/10.1037/a0039413

Pope, N. G. (2019) ‘The effect of teacher ratings on teacher performance’, *Journal of Public Economics*, 172, pp. 84–110. doi: https://doi.org/10.1016/j.jpubeco.2019.01.001.

Prelec, D., & Loewenstein, G. (1991). Decision making over time and under uncertainty: A common approach. *Management Science*, *37*(7), 770–786.

Rachlin, H., & Green, L. (1972). COMMITMENT, CHOICE AND SELF-CONTROL1. *Journal of the Experimental Analysis of Behavior*, *17*(1), 15–22. https://doi.org/10.1901/jeab.1972.17-15

Read, D. (2001) ‘Is Time-Discounting Hyperbolic or Subadditive?’, *Journal of Risk and Uncertainty*, 23(1), pp. 5–32. doi: 10.1023/A:1011198414683.

Read, D. *et al.* (2005) ‘Four Score Date / Delay and Seven Years from Now : The Effect in Temporal Discounting’, *Management Science*, 51, pp. 1326–1335. doi: 10.1287/mnsc.l050.0412.

Read, D., Frederick, S. and Scholten, M. (2012) ‘DRIFT: An Analysis of Outcome Framing in Intertemporal Choice.’, *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: 10.1037/a0029177.

Read, D. and Van Leeuwen, B. (1998) ‘Predicting hunger: The effects of appetite and delay on choice’, *Organizational behavior and human decision processes*. Elsevier, 76(2), pp. 189–205.

Read, D., Olivola, C. Y. and Hardisty, D. J. (2016) ‘The Value of Nothing: Asymmetric Attention to Opportunity Costs Drives Intertemporal Decision Making’, *Management Science*. INFORMS, 63(12), pp. 4277–4297. doi: 10.1287/mnsc.2016.2547.

Reuben, E., Sapienza, P. and Zingales, L. (2010) 'Time discounting for primary and monetary rewards', *Economics Letters*. Elsevier, 106(2), pp. 125–127. doi: 10.1016/j.econlet.2009.10.020.

Reynolds, B. and Schiffbauer, R. (2004) 'Measuring state changes in human delay discounting: an experiential discounting task.', *Behavioural processes*, 67(3), pp. 343–356. doi: 10.1016/j.beproc.2004.06.003.

Richards, J. B. *et al.* (1997) 'Determination of discount functions in rats with an adjusting-amount procedure', *Journal of the experimental analysis of behavior*. Wiley Online Library, 67(3), pp. 353–366.

Rogers, T. and Feller, A. (2016) 'Discouraged by Peer Excellence Exposure to Exemplary Peer Performance Causes Quitting', *Psychological science*. SAGE Publications, p. 0956797615623770.

Roth, Y., Wänke, M. and Erev, I. (2016) 'Click or Skip: The Role of Experience in Easy-Click Checking Decisions', *Journal of Consumer Research*, 43(4), pp. 583–597. doi: 10.1093/jcr/ucw053.

Rubinstein, A. (1988). Similarity and decision-making under risk (is there a utility theory resolution to the Allais paradox?). *Journal of Economic Theory*, *46*(1), 145–153. https://doi.org/https://doi.org/10.1016/0022-0531(88)90154-8

Samuelson, P. A. (1937) 'A note on measurement of utility', *The Review of Economic Studies*. JSTOR, 4(2), pp. 155–161.

Scholten, M. and Read, D. (2006) 'Beyond discounting: the tradeoff model of intertemporal choice'. Operational Research Group, Department of Management, London School of Economics and Political Science.

Scholten, M. and Read, D. (2010) 'The psychology of intertemporal tradeoffs.', *Psychological Review*. American Psychological Association, 117(3), p. 925.

Scholten, M. and Read, D. (2013) 'Time and outcome framing in intertemporal tradeoffs.', *Journal of Experimental Psychology: Learning, Memory, and Cognition*. American Psychological Association, 39(4), p. 1192.

Singmann, H. *et al.* (2016) 'afex: analysis of factorial experiments. R package version 0.16-1', *R package version 0.16-1*.

Solnick, J. V *et al.* (1980) 'An experimental analysis of impulsivity and impulse control in humans', *Learning and Motivation*. Elsevier, 11(1), pp. 61–77.

Spektor, M. S. *et al.* (2019) 'How similarity between choice options affects decisions from experience: The accentuation-of-differences model.', *Psychological Review*.

Stevens, S. S. (1975) *Psychophysics*. Transaction Publishers. Available at: http://books.google.co.uk/books?id=r5JOHlXX8bgC.

Stevens, S. S. and Greenbaum, H. B. (1966) 'Regression effect in psychophysical judgment', *Perception & Psychophysics*. Springer, 1(12), pp. 439–446.

Strotz, R. H. (1955) *Myopia and inconsistency in dynamic utility maximization*, *The Review of Economic Studies*. doi: 10.2307/2295722.

Takahashi, T. (2005) 'Loss of self-control in intertemporal choice may be attributable to logarithmic time-perception', *Medical hypotheses*. Elsevier, 65(4), pp. 691–693.

Takahashi, T. (2006) 'Time-estimation error following Weber–Fechner law may explain subadditive time-discounting', *Medical hypotheses*. Elsevier, 67(6), pp. 1372–1374.

Taylor, M. J. *et al.* (2015) 'Improving social norms interventions: Rank-framing increases excessive alcohol drinkers' information-seeking', *Health Psychology*. American Psychological Association, 34(12), pp. 1200–1203. doi: 10.1037/hea0000237.

Thaler, R. (1981) 'Some empirical evidence on dynamic inconsistency', *Economics Letters*, 8(3), pp. 201–207. doi: https://doi.org/10.1016/0165-1765(81)90067-7.

Trautmann, S. T., & van de Kuilen, G. (2015). Ambiguity attitudes. *The Wiley Blackwell Handbook of Judgment and Decision Making*, *1*, 89–116.

Trueblood, J. S. (2012) 'Multialternative context effects obtained using an inference task', *Psychonomic bulletin & review*. Springer, 19(5), pp. 962–968.

Trueblood, J. S. *et al.* (2013) 'Not just for consumers: Context effects are fundamental to decision making', *Psychological science*. Sage Publications Sage CA: Los Angeles, CA, 24(6), pp. 901–908.

Tsetsos, K., Chater, N. and Usher, M. (2012) 'Salience driven value integration explains decision biases and preference reversal', *Proceedings of the National Academy of Sciences*. National Acad Sciences, 109(24), pp. 9659–9664.

Tversky, A. and Kahneman, D. (1981) 'The framing of decisions and the psychology of choice', *Science*. American Association for the Advancement of Science, 211(4481), pp. 453–458.

Tversky, A. and Kahneman, D. (1992) 'Advances in prospect theory: Cumulative representation of uncertainty', *Journal of Risk and Uncertainty*, 5(4), pp. 297–323. doi: 10.1007/BF00122574.

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science*, *20*(4), 473–479.

Weber, E. U., Shafir, S. and Blais, A.-R. (2004) 'Predicting risk sensitivity in humans and lower animals: risk as variance or coefficient of variation.', *Psychological review*. American Psychological Association, 111(2), p. 430.

Wulff, D. U. and van den Bos, W. (2017) 'Modeling Choices in Delay Discounting', *Psychological Science*. SAGE Publications Inc, 29(11), pp. 1890–1894. doi: 10.1177/0956797616664342.

Wulff, D. U., Mergenthaler-Canseco, M. and Hertwig, R. (2018) 'A meta-analytic review of two modes of learning and the description-experience gap', *Psychological Bulletin*, 144(2), pp. 140–176. doi: 10.1037/bul0000115.

Zauberman, G. *et al.* (2009) 'Discounting time and time discounting: Subjective time perception and intertemporal preferences', *Journal of Marketing Research*. American Marketing Association, 46(4), pp. 543–556.