The effect of episodic future thinking on delay discounting

By

Rebecca Olsen

A thesis

submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree of

Doctor of Philosophy in Psychology

Victoria University of Wellington

2020

Acknowledgements

I am extremely grateful for the guidance and support from my supervisors: Dr. Maree Hunt and Dr. Anne Macaskill. Thank you for making the lab a fun place to be (with lots of cake) and always encouraging me to think critically. I also want to thank the other students from the Human Learning Lab for their good company and suggestions.

The greatest challenge in completing this thesis was conducting the meta-analysis, which would not have been possible without the help of many. Thank you to Dr. Lisa Woods and Professor Garth Fletcher for the statistical advice. I am also grateful to the following researchers for providing data: Dr. Roland Benoit, Dr. Adam Bulley, Dr. Joseph Kable, Dr. Karolina Lempert, Dr. Jenkin Mok, Dr. Daniela Palombo, Dr. Jan Peters and Dr. Sarah Snider. Most of all, I would like to thank Jared Pickett for the many hours he spent completing the reliability coding, and Dr. Rana Asgarova for completing a subset of the coding after Jared began his much deserved full-time job.

I also want to thank my family and friends for supporting me throughout. I am most grateful to my partner, Tom Fletcher. He has encouraged me to persevere on days when it felt impossible. My weekly visits to Nana's house also provided me with much-needed guilt-free breaks, where she patiently listened to all of my worries and assured me that everything would work out.

Finally, this thesis would not have been possible without the financial support of the Victoria University of Wellington doctoral and submission scholarships.

Abstract

Delay discounting refers to the fact that rewards lose their value if they are delayed. Excessive delay discounting is associated with various health-related problems such as over-eating and substance abuse. One phenomenon shown to reduce delay discounting is Episodic Future Thinking (EFT; imagining personal future events). Across multiple experiments and a meta-analysis, the current thesis examined the reliability of the effect of EFT on delay discounting and also sought to clarify the components of EFT that are necessary to reduce delay discounting.

Experiment 1 replicated the EFT effect using a common titrating-amount procedure, and the meta-analysis based on 40 independent studies confirmed that EFT has a reliable, medium-sized effect on delay discounting. The meta-analysis also assessed the robustness of the EFT effect across various methodological features and participant characteristics. A multiple meta-regression revealed that the between-study variability in the size of the EFT effect was accounted for by study design and type of discounting measure. Within-subjects design studies had significantly smaller effect sizes than between-subjects design studies. Studies that used *k* as a discounting rate measure had significantly smaller effect sizes than studies that used area under the discounting curve or other atheoretical discounting measures. The size of the EFT effect was robust across different participant characteristics and ages, suggesting that EFT may be utilized as an effective intervention for a variety of age groups and impulsive behaviors.

Experiments 2A, 2B and 2C assessed the suitability of an alternative and briefer delay discounting procedure (in which participants simply report their indifference points) for subsequent EFT experiments. The report indifference points procedure produced a high percentage of invalid data, a finding that was replicated across all three experiments. We therefore continued to use the well-established titrating-amount procedure in subsequent studies.

Experiment 3A found that episodic past thinking (imagining personal past events) and semantic future thinking (estimating what a person could buy with the larger, delayed amount) had no effect on discounting, indicating that both episodic thinking and future thinking are necessary components of the EFT effect. Experiment 3A results also indicated that financial relevance alone is not sufficient to reducing discounting. Experiment 3B found that the future events also need to be personally relevant to reduce discounting, and that

participants perceived EFT to reduce their discounting by primarily making the larger, delayed reward more valuable. We also showed that demand characteristics, where participants change their behavior to conform to the researcher's expectations, are an unlikely explanation for the EFT effects found in Experiments 3A and 3B. Further research is warranted to form a better understanding of the mechanism(s) through which EFT reduces delay discounting.

Table of Contents

List of Tables

List of Figures

Chapter 1: General Introduction

Delay discounting refers to the fact that rewards lose their value if they are delayed. For example, choosing between $500 now and $1,000 now is an easy choice because $1,000 is worth more than $500. Choosing between $1,000 now and $1,000 in 6 months is also an easy choice because people want rewards now rather than later. However, choosing between $500 now and $1,000 in 6 months can be a difficult choice, and many people will choose $500 now even though it is the smaller amount. This is because the $1,000 has lost value due to the delay to its receipt. An everyday example where people have to choose between delay and amount is saving to buy a house. Choosing to go out for coffee instead is the smaller, sooner reinforcer, whereas saving now would lead to the larger, delayed reinforcer of owning a house. Another example where delayed reinforcement might play a role is academic procrastination; choosing to watch television (or some other procrastination behaviour) would lead to a smaller, sooner reinforcer, whereas studying would lead to the larger, delayed reinforcer of a good grade or gaining a degree. In these types of delay-amount trade-off decisions, choosing the smaller, sooner reinforcer is characterised as the *impulsive choice*, whereas choosing the larger, delayed reinforcer is characterised as the *self-controlled choice*.

Measuring delay discounting in the laboratory

Researchers use the *hypothetical money task* as an experimental analogue of delay-amount trade-off decisions. Participants make a series of choices between a smaller, sooner reinforcer (e.g., $500 now) and a larger, delayed reinforcer (e.g., $1,000 in one year). The primary purpose of the hypothetical money task is to determine the rate at which reinforcers lose value as a function of delay. While this can be achieved using a variety of procedures (for a review, see Madden & Johnson, 2010), the *titrating-amount procedure* is commonly used (e.g., Bulley & Gullo, 2017; Du, Green, & Myerson, 2002; Fernie, Cole, Goudie, & Field, 2010; Holt, Green, & Myerson, 2003; Myerson, Green, Hansen, Holt, & Estle, 2003; Sellitto, Ciaramelli, & di Pellegrino, 2010; Snider, LaConte, & Bickel, 2016; Stein et al., 2017; Yi, Stuppy-Sullivan, Pickover, & Landes, 2017). The titrating-amount procedure is a variant of the hypothetical money task where the smaller, sooner amount of money is either increased or decreased depending on the participant's previous choice (Du et al., 2002). For example, if a participant chose '$500 now' over '$1000 in 1 year', then during the next trial for the 1 year delay the smaller, sooner amount would decrease in order to make the larger, delayed alternative more appealing. If a participant chose '$1,000 in 1 year', then during the next trial for that delay the researcher would increase the smaller, sooner amount in order to make the smaller, sooner alternative more appealing. It is then possible to determine the point at which the smaller, sooner amount is subjectively equivalent to the larger, delayed amount for a particular individual at a particular delay. This is known as an *indifference point*, which is identified for several delays. For example, if the participant chose $1,000 in 1 year over $600 now, and then on the next trial they switched their choice from '$1,000 in 1 year' to '$700 now', then they would be assigned an indifference point of $650 at the 1-year delay. That is, receiving $650 now is equal in value to the participant as receiving $1,000 in 1 year. This is because the smaller, sooner amount of the previous trial ($600) was not large enough for the participant to switch their choice, however, the smaller, sooner amount of the current trial ($700) was large enough. Thus, the middle value of the two smaller, sooner values is estimated as the point at which they are indifferent ($650). The indifference point is interpreted as the *subjective value* of the larger, delayed reward. Regardless of the procedure used, subjective value generally decreases as the delay to the larger amount increases, and subjective value decreases more quickly at delays closer to the present.

Delay discounting rate (i.e., degree of impulsivity on the hypothetical money task) has been measured in various ways within the literature. A common discounting rate measure is *k*, which is determined by using theory-driven mathematical models (e.g., the hyperbola) to describe the relationship between delay and the indifference points.

$$V = \frac{A}{1+kD}$$ [Equation 1.1; Mazur, 1987]

Equation 1.1 describes the hyperbolic model where *V* is the subjective value of the reinforcer, *A* is the undiscounted value of the reinforcer, *D* is the delay to the reinforcer, and *k* (a free parameter that is identified once the model has been fitted to the indifference points) is the rate at which that individual discounts that commodity in that context. A higher *k* indicates relatively impulsive decision making (i.e., a higher discounting rate) whereas a lower *k* indicates relatively self-controlled decision making (i.e., a lower discounting rate). The hyperbolic equation describes a larger proportional decrease in subjective value at shorter delays than at longer delays (Myerson & Green, 1995), and therefore provides a good description of delay discounting (e.g., Green & Myerson, 1996; Green, Myerson, & Ostaszewski, 1999; Madden, Begotka, Raiff, & Kastern, 2003; Madden, Bickel, & Jacobs, 1999; Myerson & Green, 1995; Rachlin, Raineri, & Cross, 1991). Figure 1.1 depicts the hyperbolic model fitted to hypothetical indifference points for illustrative purposes.

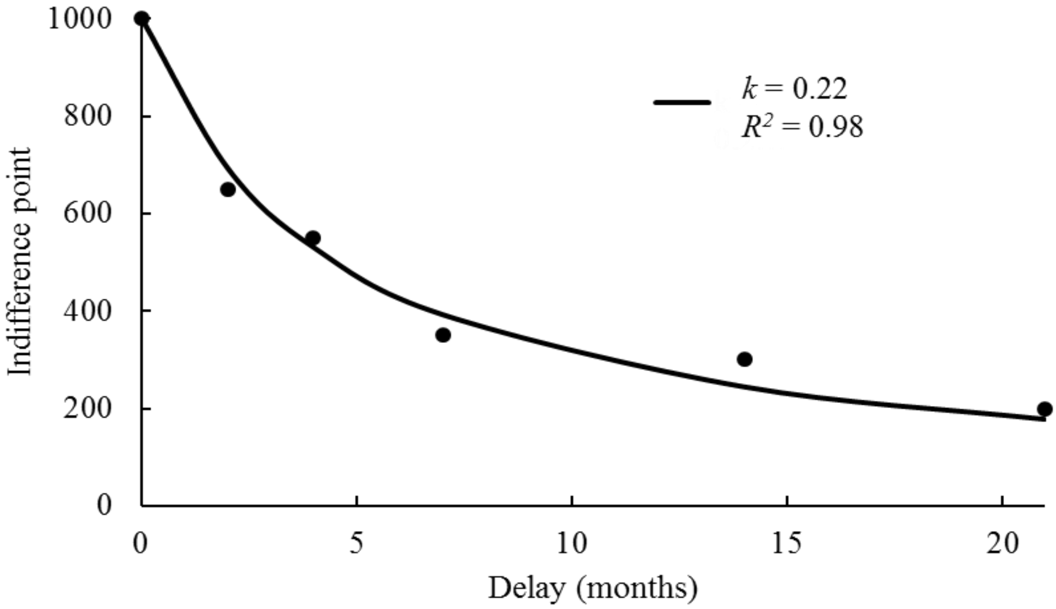

Figure 1.1. Indifference points as a function of increasing delay using hypothetical data (points). The line represents the best-fitting hyperbolic model.

Unlike the hyperbolic model, atheoretical measures do not require any assumptions regarding the mathematical form of the discounting function. Atheoretical discounting rate measures include: 1) directly analyzing the indifference points and retaining delay as an independent variable, 2) calculating the percentage/proportion of smaller, sooner (or larger, delayed) choices made (*choice index*; Benoit, Gilbert, & Burgess, 2011; Boettiger et al., 2007), 3) calculating the accumulated rewards over trials as a proportion of the maximum accumulated rewards possible (*reward index*; Benoit et al., 2011; Boettiger et al., 2007), and 4) calculating *Area Under the Curve (AUC)*. AUC is calculated as:

$$\text{AUC} = \sum x_2 - x_1 \left(\frac{y_1 + y_2}{2}\right)$$ [Equation 1.2; Myerson Green, & Warusawitharana, 2001]

where x_1 and x_2 refer to the successive delays and y_1 and y_2 refer to the subjective values associated with those delays. AUC values are typically expressed as a proportion of the total possible area. AUC values range from 0 (maximum delay discounting) to 1 (no delay discounting). The grey shading in Figure 1.2 depicts the area under the discounting curve using the same hypothetical data as shown in Figure 1.1. Accordingly, higher AUC values correspond to a greater proportion of the shaded area relative to the total area and indicate shallower discounting (note that this is in the reverse direction to *k* values in which a higher *k* corresponds to steeper discounting). In contrast to the other atheoretical measures, AUC provides a measure of discounting which takes the value of all of the delays into account.

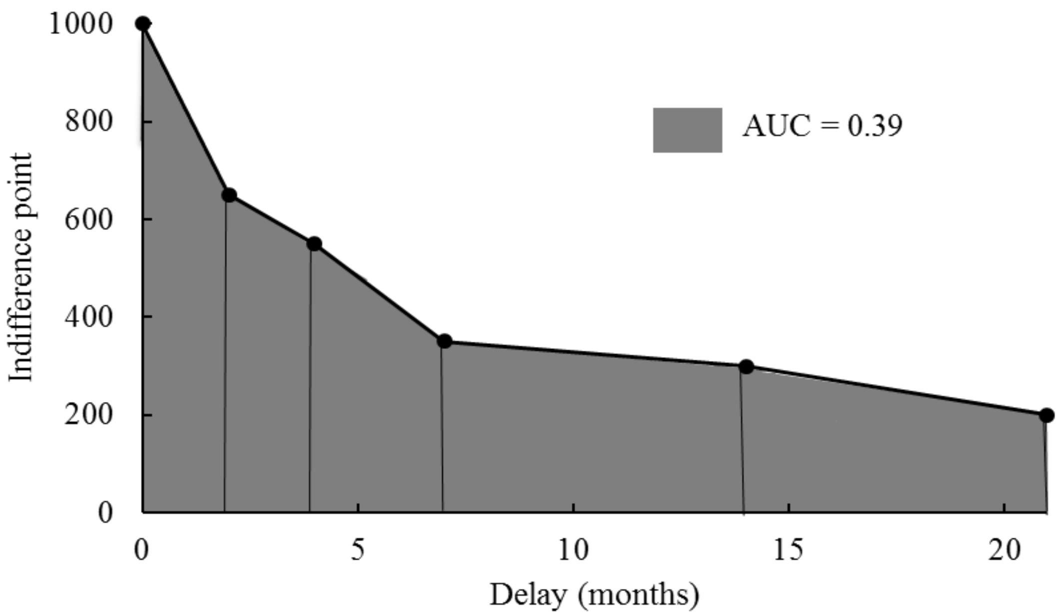

Figure 1.2. Indifference points as a function of increasing delay using the same hypothetical data as in Figure 1.1 (points). The vertical lines indicate the trapezoid areas used to calculate the area under the discounting curve (AUC). The grey shading corresponds to the total AUC.

The effects of experimental manipulations on delay discounting rate can differ depending on whether *k* or AUC is used (Olsen, Macaskill, & Hunt, 2018; Smith & Hantula, 2008). For example, Smith and Hantula (2008) found that *k* and AUC measures supported different conclusions about the combined effects of using different reinforcer magnitudes and discounting procedures. Additionally, delay discounting rate was correlated with a self-report measure of impulsiveness according to *k*, but not according to AUC in this study. The authors concluded that it is important to consider both *k* and AUC, in order to form a better understanding of discounting processes. Chapters 3 and 5 discuss potential reasons for differing results between *k* and AUC within the context of the current thesis findings.

Research into delay discounting provides insight into a range of impulsive behaviors, such as over-eating (Amlung, Petker, Jackson, Balodis & MacKillop, 2016), problem gambling (Callan, Shead, & Olson, 2011; Petry & Madden, 2010; Reynolds, 2006), drug abuse (Bickel & Marsh, 2001; De Wit, 2009; Perry & Carol, 2008; Washio et al., 2011) and academic procrastination (Cassidy & Kangas, 2014, Olsen, 2016; Olsen et al., 2018). These behaviours may be conceptualised as choosing smaller, sooner rewards over larger, delayed rewards. For example, a cocaine-dependent individual may be viewed as choosing the immediate effects of cocaine over larger, delayed rewards associated with abstinence such as improved health and family relationships. Individuals with obesity, problem gambling disorder, drug-dependency and high academic procrastination tendencies discount delayed rewards more steeply than matched controls (Amlung et al., 2016; Cassidy & Kangas, 2014; Davis, Patte, Curtis, & Reid, 2010; Dixon, Marley, & Jacobs, 2003; Madden, Petry, Badger, & Bickel, 1997; Mitchell, 1999; Olsen, 2016; Petry, 2001, Weller, Cook, Avsar, & Cox, 2008), indicating that discounting rates are a valid way to conceptualise impulsivity.

Furthermore, Washio et al. (2011) found that people recovering from cocaine addiction who were more impulsive on the hypothetical money task were likely to relapse sooner than recovering cocaine addicts who were less impulsive on the money task. Higher delay discounting rates have also been shown to predict sooner relapse in cigarette smokers (Dallery & Raif, 2007; Krishnan-Sarin et al., 2007; Yoon et al., 2007), and lower grade point averages in university students (Kirby, Winston & Santiesteban, 2005). Given the prevalence of delay-amount trade-off decisions in our everyday lives, and relationships between delay discounting and behaviors of applied relevance, investigating ways to reduce delay discounting rate would be beneficial.

The effect of episodic future thinking (EFT) on delay discounting rate

One phenomenon shown to decrease delay discounting rate is Episodic Future Thinking (EFT; e.g., Benoit et al., 2011; Peters & Büchel, 2010; see Chapter 3 of the current thesis [meta-analysis] for 36 additional citations). EFT is the act of imagining personal future events, which involves *episodic* simulation (i.e., thinking about an event with an entire scene unfolding), rather than *semantic* thinking (generalized knowledge without a scene unfolding; Atance & O'Neill, 2001). Generally, participants are first asked to write about positive future events (including what they did, where they were, who they were with, accompanying feelings) and then they are cued to think about these events before making choices during the hypothetical money task.

Peters and Büchel (2010) were one of the first to investigate the effect of EFT on delay discounting rate. Participants listed positive future events that they had planned for the next seven months, then completed a delay discounting task where half of the delays matched the delays of previously listed events (EFT delays), and the other half of the delays did not (control delays). All delays were presented in a random order for each participant. During the EFT delays, cues for the imagined future events were presented below the larger, delayed choices, which was intended to evoke EFT. For example, if a participant said that they were going on a vacation in Paris in 45 days, then the cue phrase 'vacation Paris' would appear under the larger, delayed option at the 45-day delay during the money task. In contrast to other EFT studies, Peters and Büchel did not explicitly instruct their participants to imagine the events associated with the cues during the money task. After the money task, participants rated how often each cue made them think about the associated imagined event during the money task, and how vivid the imagined event was. These ratings were averaged across events and the two dimensions to calculate an episodic imagery score. Money discounting rates were significantly lower for the EFT delays compared to the control delays, and the size of the difference in discounting rate between the two delay types was positively related to degree of episodic imagery. EFT also produced greater brain activation in the prefrontal cortex and hippocampus, which have been previously implicated in cognitive control and construction of mental imagery, respectively (Schacter, Addis, & Buckner, 2007). The latter finding is consistent with the conclusion that participants did imagine the events associated with the cues during the money task, even though they were not explicitly instructed to do so. The authors concluded that EFT reduced delay discounting rate.

The effect of EFT on impulsive behaviours

There is emerging evidence that EFT also reduces problematic behaviours related to steep discounting. For example, Stein et al. (2016) investigated the effects of EFT on both delay discounting rate and cigarette self-administration in nicotine-dependent individuals. Participants were randomly allocated to either an EFT condition or a control condition. The EFT group were instructed to generate positive future events that could realistically occur at five particular time points. In contrast, the control group were instructed to report real positive events that occurred during the previous day at five particular time points. Participants then completed a monetary delay discounting task where they were cued to think about their previously generated future or recent events. Lastly, participants completed a cigarette self-administration task during which they were cued to think about these same events. Participants in the EFT group had lower discounting rates and smoked less than participants in the control group. Importantly, participants in both groups were told to generate events that were unrelated to smoking, meaning that the future events do not need to be related to the current decision.

Within laboratory settings, researchers have also found that EFT reduces caloric intake in obese/overweight individuals (Daniel, Stanton, & Epstein 2013b), and alcohol demand in alcohol-dependent individuals (Snider et al., 2016). Outside of the laboratory, EFT has also been shown to reduce, not only caloric intake within a food-court setting (O'Neill, Daniel, & Epstein, 2016), but also body mass index in overweight individuals (Sze, Daniel, Kilanowski, Collins, & Epstein, 2015), further highlighting the therapeutic potential of EFT. Thus, although little research has investigated the effects of EFT on problematic behaviours outside of the laboratory, these promising results encourage further laboratory-based research to gain a better understanding of the circumstances under which EFT increases self-control.

Which components of EFT contribute to reduced delay discounting rate?

When participants are asked to engage in EFT, they are typically asked to consider events that are emotionally positive, plausible, personally relevant, and in the future. They are also typically asked to consider where they are, who they are with and accompanying feelings, in order to encourage episodic thinking. Thus, EFT consists of multiple, separable components. It is important to understand which components of EFT are necessary (i.e., without which there is no effect) and sufficient (i.e., the only thing needed to get the effect) to reduce discounting. It is also important to understand which components are contributors to the EFT effect (i.e., not necessary nor sufficient, but alter the size of the effect). This will result in EFT being translated into clinical interventions more efficiently. Below we review the importance of the following components based on the current literature: positive emotional valence, plausibility, financial relevance, episodic thinking and future thinking.

Positive emotional valence. Most researchers have assessed the effects of emotionally-positive future events on delay discounting rate. There are, however, a few studies that have assessed the effects of negative and neutral future events on delay discounting rate. Both Liu, Feng, Chen, and Li (2013) and Zhang, Peng, Qin, Suo, and Feng (2018) found that, while thinking about positive future events reduced discounting rate, thinking about negative future events *increased* discounting rate, and thinking about neutral future events had no effect. These results indicate that the positive emotional valence component of EFT may be necessary to reduce delay discounting rate.

These results raised the possibility that EFT reduces delay discounting through increasing positive mood. However, the effect of mood inductions alone on delay discounting differs from the effects of positive EFT and negative EFT, indicating that the effect of EFT on delay discounting cannot be explained as a simple mood induction effect. For example, Hirsh, Guindon, Morisano and Peterson (2010) investigated the effects of both positive and negative mood inductions on delay discounting rate. Participants rated their current positive and negative mood, and then mood was manipulated by having participants complete a puzzle either before a confederate completed it ("participant success"/positive mood condition) or after a confederate completed it ("participant failure"/negative mood condition). Participants then rated their positive and negative mood again, followed by completing a delay discounting task. Positive mood was associated with increased delay discounting rate, but only in extraverted individuals. Negative mood was not associated with delay discounting rate. Consistent with Hirsh et al. (2010), several other studies have also shown that viewing positive affective images (including erotic stimuli, attractive faces or desserts) before a delay discounting task increases discounting rate (Kim & Zauberman, 2013; Li, 2007; Van den Bergh, Dewitte, & Warlop, 2008; Wilson & Daly, 2004).

Owens (2015) also investigated the effects of negative and neutral mood inductions on delay discounting rate. In this study, mood was manipulated by instructing participants to read established mood induction statements prior to completing the discounting task. Mood-condition-congruent music was also played throughout the session. Consistent with Hirsh et al. (2010), Owens found that the negative mood induction had no effect on delay discounting rate. Additionally, the neutral mood induction had no effect on delay discounting rate.

In summary, although positive emotional valence may be a necessary component of the EFT effect (Liu et al., 2013; Zhang et al., 2018), increasing positive mood is not the mechanism through which EFT reduces discounting. There is also additional evidence that components of EFT other than positive emotion are also important contributors to the effect (as outlined below).

Plausibility. Researchers that have assessed the effects of EFT on delay discounting rate have instructed participants to imagine plausible future events. Although no study has compared the effects of EFT on delay discounting rate when future events are plausible vs. non-plausible, the literature on mental contrasting and positive fantasies indicates that the EFT effect would likely be weaker if researchers instructed participants to imagine implausible future events. Findings from this separate literature demonstrate that people are more likely to work towards future goals if they are plausible. Thus, the plausibility of imagined future events might be important in helping people obtain the larger, delayed reward in delay-amount trade-off situations.

According to fantasy realization theory (Oettingen, 1996; 1999), *mental contrasting* involves linking desires with reality whereby people first elaborate on a desired future and then on aspects of the present reality (i.e., obstacles that could prevent that desired future). This mental contrasting process causes people to realise that they must act to obtain their desired future, and therefore allows people to judge the plausibility of their future goals. How likely people perceive their desired future to be will then determine future goal commitments. That is, if people judge the desired future as likely, then this results in strong goal commitments. This is consistent with the effect of EFT on discounting, as instructing participants to imagine positive, plausible future events makes people more likely to choose the larger, future reward. However, fantasy realization theory also states that if people judge the desired future as unlikely, then this results in weak goal commitments. Additionally, mental contrasting differs from *indulging* where people only imagine the desired future (i.e., without considering the plausibility of that desired future) and *dwelling* where people only consider the negative reality (Oettingen & Gollwitzer, 2009). If people only fantasize (indulge) about the desired future or only consider the present reality (dwell), then this results in only moderate goal commitments that are independent of likelihood judgements. Thus, imagining an implausible future event (or simply not considering the plausibility of a future event) may result in a weaker commitment to choosing the larger, delayed reward.

Several studies have tested the effects of mental contrasting, indulging and dwelling on future goal attainment and have shown similar results that are consistent with fantasy realization theory (see Oettingen & Gollwitzer, 2009 for a review). For example, Oettingen, Mayer and Thorpe (2010) compared the effects of mental contrasting, indulging and dwelling on commitment to reduce cigarette consumption. Participants rated how likely they were to reduce their cigarette consumption (i.e., their expectations of success). Participants in the mental contrasting group then wrote about four aspects associated with a positive future of reducing or quitting smoking (e.g., more energy, better skin), followed by four aspects of reality that could prevent this positive future (e.g., stress, peer pressure). Participants in the indulging group only wrote about the positive future aspects, and participants in the dwelling group only wrote about the negative reality aspects. Two weeks later participants were mailed a questionnaire where they had to report the exact date they had taken their most difficult step in order to reduce their cigarette consumption. Participants in the mental contrasting group reported taking steps to reduce cigarette consumption sooner than participants in the other two groups, but only when their expectations of success were high. When expectations of success were low, participants in the mental contrasting group reported taking steps later than participants in the other two groups. Immediacy of action was the same between the indulging and dwelling groups, regardless of expectations of success. Overall, these results suggest that people are more likely to commit to future goals when they consider the plausibility of the desired future, and they judge the future outcome as likely.

Studies have also found that judging the desired future outcome as likely predicted more weight loss in obese individuals (Oettingen & Wadden, 1991) and higher grades in university students (Oettingen & Mayer, 2002) than positive fantasies regarding the future outcome alone. Although the above studies did not measure delay discounting, they did assess behaviours that are related to delay discounting (i.e., cigarette smoking, Bickel, Odum & Madden, 1999; academic performance, Kirby et al., 2005; weight loss, Amlung et al., 2016), indicating that plausibility may be an important contributor to the effect of EFT on delay discounting.

Financial relevance. Another potential component of the EFT effect is financial relevance. Although most researchers do not explicitly tell participants to imagine spending the larger, delayed reward on the future event (although see Benoit et al., 2011; Palombo, Keane, & Verfaellie, 2015), participants could still be thinking about the cost of the future events they are instructed to think about when making decisions during the money task (given that the future events could be improved by having more money available to spend on them). O'Donnell, Daniel, and Epstein (2017) compared the effects of financially-goal related EFT (e.g., ''In 2 weeks I am purchasing a new computer") and general EFT (e.g., ''In 2 weeks I am going home for the weekend") on delay discounting rate. Participants were instructed to describe where they were, how they were feeling, who they were with and what they were doing, to encourage episodic thinking in both EFT conditions. Both EFT conditions reduced discounting rate compared to control, but the effect was larger for financially-goal related EFT. These results show that, while the financial-relevance component of EFT may contribute to the effect, it is not *necessary* to produce the effect. Thus, EFT effects are more pronounced in, but are not limited to, situations when the event is related to the current decision (i.e., when both require monetary considerations). These results are also consistent with Stein et al. (2016), who found that imagining future events that were unrelated to smoking reduced cigarette self-administration in nicotine-dependent individuals, and thus further demonstrate that the future events do not need to be related to current decision [1].

Episodic and future thinking. Similar EFT conditions to that of Peters and Büchel (2010) have been used in other studies investigating the effect of EFT on delay discounting rate, however the control conditions used have differed widely. These different control conditions have been used to isolate the importance of the episodic and future thinking components of EFT.

Benoit et al. (2011) used a control condition that allowed them to isolate the importance of episodic thinking in the EFT effect. Participants imagined episodes of spending money for a range of scenarios (EFT condition) or estimated what the money could purchase in the scenarios (semantic future thinking control condition), before making decisions in the money discounting task. Unlike Peter and Büchel's (2010) study, the EFT and control conditions shared the same semantic retrieval demands. That is, both conditions required the same amount of knowledge about goods and prices, and therefore semantic thought can be ruled out as an explanation for any differences in discounting rate obtained between the two conditions. Consistent with Peter and Büchel's findings, discounting rates were significantly lower for the EFT condition compared to the control condition, and in the EFT condition there was greater activity in brain regions which have been previously implicated in cognitive control (prefrontal cortex) and construction of mental imagery (hippocampus).

Although Benoit et al.'s (2011) methodology controls for semantic retrieval demands, their control ('estimate') task is a non-episodic future thinking task. Different control conditions used by other researchers involved instructing participants to think about events that have already happened, which ensures that both conditions elicit episodic thinking, and therefore isolates the importance of future thinking in the EFT effect (Daniel, Stanton, & Epstein, 2013a; Daniel et al., 2013b; Dassen et al., 2016; Stein et al., 2016). For example, Dassen et al. (2016) instructed participants to either imagine events that they have planned in the future at given delays (one week, two weeks, one month, six months; EFT condition) or to recall events that have recently happened (one day ago, two days ago, a week ago, a month

[1] Although see Dassen, Jansen, Nederkoorn and Houben (2016) who found that while both food-related EFT and non-food-related EFT reduced discounting, only food-related EFT reduced caloric intake.

ago), which they called the Episodic Past Thinking (EPT) condition. Delays used in the subsequent money discounting task were similar to the delays used in the EFT manipulation, and participants were told to imagine the events they just wrote about before making decisions. Participants in the EFT condition had significantly lower discounting rates compared to those in the EPT condition, and there were no group differences in how vivid, positive or detailed the events were.

The studies described above demonstrate that two components of EFT are important: episodic thinking and future thinking. That is, findings that EFT reduces delay discounting rate compared to semantic future thinking (e.g., Benoit et al., 2011) show that episodic thinking is more effective at reducing delay discounting rate than semantic thinking. Additionally, findings that EFT reduces delay discounting rate compared to EPT (e.g., Dassen et al., 2016) shows that thinking about the future is more effective at reducing delay discounting rate than thinking about the past. However, it remains unclear whether one of these components is more important than the other, and whether each of these components are *necessary* to produce the EFT effect (see introduction to Experiment 3A for a more comprehensive review).

Demand characteristics

To reiterate, in a typical EFT procedure, participants are asked to imagine future events at specific points in time and to generate cue words for those events. The cue word is then presented on trials during the delay discounting task that match that delay. For example, if a participant previously wrote about attending a friend's wedding in one year, then the cue word 'wedding' would be presented when they made a choice between a smaller, sooner amount of money and a larger amount of money to be received in one year. These cues may reveal the researcher's hypothesis that EFT will increase preference for the larger, delayed reward. Therefore, demand characteristics, where participants change their behavior to conform to the researcher's expectations (Orne, 1962), are a concern within EFT procedures (see Rung & Madden, 2018a for a discussion).

A demand characteristics explanation seems particularly likely in studies where delays in the discounting task match the delays in the EFT manipulation, but not the delays in the control condition (e.g., as in Dassen et al., 2016, as described above). However, concerns over the content of EFT matching content within the delay discounting task have been somewhat mitigated by research by O'Donnell, Hollis-Hansen, and Epstein (2019). These researchers investigated whether the time-points of EFT cues needed to match the delays in the delay discounting task in order to be effective. Participants in the 'matched' group listed five future events at particular time-points (1 day, 1 week, 1 month, 6 months, 2 years), and were subsequently cued to imagine particular events on trials that matched the delay of the discounting task, as in typical EFT studies. Participants in the 'unmatched' group were required to list five future events at the same time-points as the 'matched group', however the time-point corresponding to the cue mismatched the delay in the discounting task. EFT groups had lower delay discounting rates compared to control groups (where participants thought about recent events), and there was no difference in the size of the EFT effect between the 'matched' and 'unmatched' groups. These results indicate that the fact that the larger, delayed amount will be received on the same date as the cued event does not necessarily encourage demand characteristics and subsequently inflate EFT effects. However, it is still important to note that the *range* of delays was still the same for the EFT cues and the discounting task. Nevertheless, in Benoit et al.'s (2011) study, EFT and the control condition (semantic future thinking) had comparable overlap in content with the discounting task (that is, for both conditions, the delays and amounts of money corresponding to the cues matched the delays and amounts of money in the discounting task), yet EFT still reduced delay discounting rate. See Experiment 3A for further discussion regarding control conditions and demand characteristics.

Further research by Stein, Tegge, Turner, and Bickel (2018) suggested that demand characteristics might partly contribute to the effect of EFT on delay discounting rate. Participants were assigned to either an EFT condition or a control condition (where they had to imagine recent events), before completing a delay discounting task. Demand was measured by asking participants what they thought the purpose of the experiment was, whether they thought they had been assigned to the experimental condition or the control condition, and their general attitudes towards the experiment. Although the effect of EFT on delay discounting rate remained significant after controlling for measures of demand characteristics, the effect did decrease. These results suggest that, while demand characteristics cannot completely account for the EFT effect, they may contribute.

Evidence that EFT can also alter behaviours that are less likely to be affected by demand characteristics (relative to hypothetical decisions in an experiment) also suggests that the EFT effect is not wholly driven by demand characteristics. That is, it is easier for a participant to help the researcher by selecting the larger, delayed reward than by reducing their smoking. Thus, EFT-reduced smoking is less likely to be a result of demand characteristics. The effect of EFT on cigarette smoking found in Stein et al.'s (2016) study is also unlikely to be due to demand characteristics given that participants were instructed to imagine events that were unrelated to smoking.

Potential mechanisms of the EFT effect

As described above, while demand characteristics may contribute to EFT effects to an extent, evidence to date suggests that they cannot completely account for the effect (Stein et al., 2018). Another possibility is that the EFT effect may be explained as a date-delay framing effect. The date-delay framing effect refers to the fact that framing delays as specific dates (e.g., $1,000 on 31st October 2020), rather than as delays (e.g., $1,000 in 1 year) reduces discounting rate. Given that EFT involves thinking about future events at a particular point in time, it is possible that EFT may reduce delay discounting because it encourages people to conceptualize the delays as dates. EFT may also reduce discounting by increasing the perceived certainty that the larger, delayed reward will be received, as it is well known that people discount rewards that are uncertain (this is known as probability discounting; see Green & Myerson, 2010 for a review). Peters and Büchel (2010) investigated these two potential mechanisms by having participants rate: 1) whether they focused on the delays as presented during the money task or whether they converted them into dates, and 2) how confident they were that they would receive their rewards. There was no significant difference in date/delay ratings or reward confidence ratings between the EFT and control conditions. These results show that neither date-delay framing effects nor reward certainty effects can account for the effect of EFT on delay discounting rate. Below we outline other potential mechanisms within the context of relevant theories.

Construal level theory. One proposed mechanism through which EFT might reduce delay discounting is that it makes the future reward more concrete (Kim, Schnall, & White, 2013; Yi et al., 2017). Construal level theory (Trope & Liberman, 2003) proposes that people construe information on a continuum from concrete (detailed, contextualized features) to abstract (general, decontextualized features). The way information is construed can depend on temporal distance; the closer something is in time, the more concretely it will be thought of, whereas the more distant something is in time, then the more abstractly it will be thought of. People may therefore discount delayed rewards because they prefer concrete rewards (Yi et al., 2017). Thus, EFT may be described as a motivating operation, which is an event that alters the value of a reinforcer. Motivating operations can be further broken down into establishing operations (events that increase the effectiveness of a given consequence) and abolishing operations (events that decrease the effectiveness of a given consequence) (Laraway, Snycerski, Michael, & Poling, 2003). If EFT reduces discounting by making the larger, delayed reward more valuable (by making the larger, delayed reward more concrete), then EFT is an establishing operation. EFT may also reduce discounting by eliminating the abolishing operation of delay by mentally bringing the future reward to the present.

Alternatively, people may discount delayed rewards because a mismatch in construal between smaller, sooner rewards (which are construed more concretely) and larger, delayed rewards (which are construed more abstractly) makes it difficult to easily compare the two outcomes (Kim et al., 2013). That is, people may make maladaptive choices (i.e., choose the smaller, sooner reward) if they are unable to process the two choice outcomes in a fluent manner, not necessarily because they prefer concrete rewards. Thus, EFT may reduce delay discounting by aligning the level on which the smaller, sooner and larger, delayed rewards are construed. Specifically, EFT might make the larger, delayed reward *as concrete as* the smaller, sooner reward. Consistent with this explanation, Kim et al. (2013) demonstrated that manipulating how concretely or abstractly participants viewed the smaller, sooner and larger, delayed rewards affected discounting rates. The researchers provided participants with concrete details about taking a trip to Paris using both the smaller, sooner and larger, delayed rewards. In the control condition, participants were also told that the smaller, sooner and larger, delayed rewards could be used for a trip to Paris, but were not given any concrete details regarding the trip. Discounting was significantly lower in the experimental condition than in the control condition. Kim et al.'s experimental condition is similar to an EFT condition used in two studies where researchers explicitly told participants to imagine spending the larger, delayed reward on a future event (Benoit et al., 2011; Palombo et al, 2015). However, Kim et al.'s experimental condition differs from the EFT studies in that the participants also imagined spending the smaller, sooner reward on the future event. Additionally, Kim et al.'s control condition also involved associating the rewards with a future event. The key difference between the experimental and control conditions was whether participants were given any concrete details regarding the event. Kim et al. therefore concluded that making the smaller, sooner and larger, delayed rewards share a similar level of construal reduced delay discounting.

In a second study, Kim et al. (2013) found that making the smaller, sooner reward more abstract also reduced discounting. Construal level theory states that, not only does the construal of information depend on temporal distance, but also on social distance. That is, the more similar someone is to us, the more socially close they seem and are therefore construed more concretely. In contrast, the more dissimilar someone is to us, the more abstractly they will be construed (Trope & Liberman, 2003; Liviatan, Trope & Liberman, 2008). Kim et al. therefore made the smaller, sooner reward as abstract as the larger, delayed reward by instructing participants to make decisions for a stranger during the money task. In a control condition, participants were instructed to make decisions during the money task for a close friend (where the smaller, sooner reward would be more concrete than the larger, delayed reward). Participants discounted delayed rewards more shallowly for strangers than for close friends, consistent with findings from other studies (Albrecht, Volz, Sutter, Laibson, & von Cramon, 2010; Ziegler & Tunney, 2012; also see Bialaszek, Bakun, McGoun, & Zielonka, 2016; O'Connell, Christakou, Haffey, & Chakrabarti, 2013). These findings further support the conclusion that making the smaller, sooner and larger, delayed rewards share a similar level of construal reduces delay discounting.

In summary, EFT likely makes the larger, delayed reward more concrete. EFT may therefore reduce delay discounting because people prefer concrete rewards (Yi et al., 2017) or because people are more likely to make adaptive choices (i.e., choose the larger, delayed reward) when the smaller, sooner and larger, delayed rewards are construed on a similar level (Kim et al., 2013). Regardless, construal level theory suggests that EFT reduces discounting by making the larger, delayed reward more concrete.

The temporal attention hypothesis. The temporal attention hypothesis (Radu, Yi, Bickel, Gross, & McClure, 2011) suggests that delay discounting rate may be reduced by shifting attention away from “now” (i.e., away from the smaller, sooner reward). Thus, EFT may reduce delay discounting because engaging in future thought shifts attention away from the present. This suggests that thinking about the past may also be sufficient to reduce delay discounting rate. As previously noted, studies have shown that thinking about the future is more effective at reducing delay discounting than thinking about the past (e.g. Dassen et al, 2016. O’Donnell et al., 2017, Stein et al., 2017). However, these studies do not tell us whether EPT has an effect on discounting compared to a standard discounting task (see Experiment 3A for a more comprehensive review).

The prediction that EPT might also reduce discounting is also consistent with the phenomenon of preference reversal, in which preferences shift from the smaller to the larger reward when equal delays are added to both rewards (Green, Fristoe, & Myerson, 1994). For example, the smaller reward may be preferred when choosing between $500 now and $1,000 in 4 months, however the larger reward may be preferred when choosing between $500 in 6 months and $1,000 in 10 months. Another example of preference reversal is when a person sets their alarm at night with the intention of getting up and going to the gym in the morning. At this time the person has chosen the larger, delayed reward of better health and fitness in the long-run (which is obtained by going to the gym). Additionally, equal delays are added to the smaller, sooner reward (sleeping in in the morning) and the larger, delayed reward, given that the person will not be faced with their actual choice until morning. However, come morning when only the larger reward is delayed, the person might switch their preference and decide to sleep in instead of going to the gym. Thus, people are more likely to choose the larger, delayed reward when they are further in time from their choice point. This is consistent with the idea that shifting attention away from the present may broaden an individual's time frame over which they consider the consequences of their behavior, which in turn could enhance self-control. Indeed, researchers have theorized that EFT may reduce discounting by broadening temporal horizon (Lin & Epstein, 2014; Snider et al., 2016). The phenomenon of preference reversal may also be explained by construal level theory, as adding equal delays to both rewards would make both of them abstract, and so the larger reward is more likely to be preferred.

Summary and overview of experiments

In summary, EFT may be used as an effective intervention to reduce impulsive decision making, given that it has been shown to reduce delay discounting rate (e.g., Benoit et al., 2011; Peters & Büchel, 2010) and impulsive behaviours that are related to steep delay discounting (Daniel et al., 2013b; O'Neill et al., 2016; Stein et al., 2016; Sze et al., 2015). Evidence to date indicates that positive emotional valence (Liu et al., 2013; Zhang et al., 2018) and plausibility (Oettingen & Mayer, 2002; Oettingen, et al., 2010; Oettingen & Wadden, 1991) are likely important components of EFT. Financial relevance is also an important component, however it is not a necessary component (O'Donnell et al., 2017). Episodic thinking and future thinking are also important components (Benoit et al., 2011; Dassen et al., 2016), however it remains unclear whether one of these components is more important than the other, and whether each of these components are *necessary* to produce the EFT effect.

The aims of the current thesis were four-fold. Firstly, we aimed to assess how reliable/robust the EFT effect is by: 1) replicating the EFT effect (Experiment 1), and 2) conducting a meta-analysis of the effect of EFT on delay discounting rate.

Secondly, across three experiments (Experiments 2A, 2B and 2C), we aimed to assess the suitability of an alternative delay discounting procedure (in which participants simply report their indifference points; Johnson, Herrmann, & Johnson, 2015) for subsequent experiments of the current thesis. In contrast to the established titrating-amount procedure, the ‘report indifference points’ procedure is faster to complete, however it has not been used frequently in the literature. A more time-efficient procedure would allow researchers to compare more conditions within subjects and thus more effectively isolate the necessary and sufficient components of the EFT effect.

Thirdly, we aimed to determine the necessary components of EFT to reduce delay discounting rate. Experiment 3A assessed the relative importance of the episodic and future thinking components of EFT, by comparing the effects of EFT, semantic future thinking, and EPT on delay discounting rate, relative to a control condition where participants completed a standard delay discounting task. We reasoned that EPT could reduce delay discounting (meaning that the future component of EFT may not be necessary) based on the Temporal Attention Hypothesis. The meta-analysis also aimed to assess the relative importance of the episodic and future thinking components by assessing type of control condition as a moderator of the size of the EFT effect. Experiment 3B assessed whether future events have to be personally relevant in order to reduce delay discounting, as no study to date has assessed the importance of this component.

Fourthly, we aimed to better understand the mechanism(s) of the effect of EFT on delay discounting rate. To assess whether EFT likely reduces delay discounting rate by increasing the perceived value of the larger, delayed reward and/or by decreasing the perceived delay to the larger, delayed reward, we asked participants a series of questions about their choices they made during Experiment 3B. Finally, we examined the extent to which demand characteristics can account for the EFT effects found in Experiments 3A and 3B.

Chapter 2: Replicating the general EFT effect (Experiment 1)

The aim of Experiment 1 was, firstly, to replicate the general EFT effect. We therefore used methodology similar to that of the original study conducted by Peters and Büchel (2010). Participants wrote about positive future events that they had planned for six specific delays, and then completed a delay discounting task where half of the delays matched the delays of previously listed events (EFT delays), and the other half of the delays did not (control delays). During the EFT delays, cues for the imagined future events were presented below the larger, delayed choices, and participants were instructed to vividly imagine the future events associated with the cues prior to making their choices. The explicit instruction to imagine the events during the money task differs from Peters and Büchel but resembles other EFT studies (e.g., Bulley & Gullo, 2017; Daniel et al., 2013b; Lin & Epstein, 2014).

In Benoit et al.'s (2011) EFT study, participants also completed a scale assessing how much, in general, they consider the future consequences of their choices (Consideration of Future Consequences (CFC) Scale; Strathman, Gleicher, Boninger, & Edwards. 1994). They found a significant negative correlation between CFC scale scores and the size of the difference between the EFT and control conditions, meaning that the EFT effect was greatest for those who were more indifferent towards the future consequences of their actions. We therefore included the CFC scale in our study to determine if we could also replicate this finding.

Method

Participants

Participants were 39 first-year psychology students at Victoria University of Wellington. Participants were recruited via the Introduction to Psychology Research Programme (IPRP) in partial fulfilment of a course requirement. Consent was obtained prior to participation. All aspects of this experiment and all subsequent experiments in this thesis were approved by the School of Psychology Human Ethics Committee under delegated authority of the Victoria University Human Ethics Committee.

Materials

The following tasks were presented to participants on a computer using Microsoft Visual Basic Software.

EFT manipulation. Participants read the following instructions:

> We want you to imagine events that you have already planned in the future or that could realistically happen at the specified point in the future. Please describe one event for each of the delays. Pick an event that you can VIVIDLY imagine, that is emotionally POSITIVE, PERSONALLY RELEVANT and EXCITING. An example could be an upcoming holiday, a friend's wedding, or going to a movie that you have been looking forward to seeing.
>
> These images will be used in the next exercise. For each event, describe the situation below. There is also a box for you to enter a cue word or phrase for each event, for example, "wedding" or "movie."
>
> Type your description of an imagined event in X delay in the box below. Consider as many details of this event as possible (what do you do, where are you, who are you with, accompanying feelings).

Participants were then asked to select the appropriate date when the event will or could happen on a calendar. Dates on the calendar were constrained within the range of the given delay. The given delays were 0-7 days, about 2 weeks, about a month, 3-4 months, 6-8 months, 1 year.

Hypothetical money task. Participants read the following instructions:

> You will now be asked a set of hypothetical questions that require you to make choices between different amounts of money. Please read each question carefully as the amount of money you can receive and when you can receive the money will vary. There is no right or wrong answer, select the option you would be most likely to choose.
>
> During some of the scenarios, the cue words you generated for your imagined events will be presented. You do not have to base your choice on the imagined event; YOU JUST HAVE TO VIVIDLY IMAGINE THAT THIS EVENT IS REALLY HAPPENING, and next you can choose your preferred reward.

Participants were asked to choose between two hypothetical choices: one reward that was smaller in magnitude but available sooner (receive $500 today) and a larger reward

($1,000) available after some delay. There were a total of 12 delays (six for the EFT condition, six for the control condition). The delays that participants had previously selected on the calendar were used for the EFT condition. The delays for the control condition were interspersed between the delays for the EFT condition. That is, a control delay was calculated by taking the average of the two EFT delays that the control delay was between. For example, if the first EFT delay was six days and the second EFT delay was 14 days, then the first control delay would be 10 days. The sixth control delay was calculated by taking the average difference between the EFT delays and adding this to the sixth EFT delay. The order of the delays was randomised. Indifference points for each delay were determined by adjusting the size of the immediate outcome ($500 today) based on what participants chose during the previous trial. Adjustments were made in $100 increments. Thus, if a participant chose to receive $500 today, then that amount of money decreased by $100 when presented during the next trial. If a participant chose to receive $1,000 after a delay, then the '$500 today' option increased by $100 when presented during the next trial. There were five trials for each delay, and a 500ms inter-trial interval between making a choice and moving to the next choice. Indifference points for each delay were calculated by averaging the smaller, sooner amount of the final trial at that delay and the predicted smaller, sooner amount of the next trial. Cues were written in red below the larger, delayed option during EFT trials. For example, if a participant had generated the cue "wedding" at a delay of 1 year during the EFT manipulation, then they would be asked to choose between "receive $500 now" or "receive $1,000 in 1 year (wedding)." No such cues were presented during control trials.

Cue ratings. Participants rated how often each cue made them think about the associated imagined event during the money task (1 = never, 7 = always), how vivid (1 = not vivid at all, 7 = highly vivid), positive (1 = not positive at all, 7 = highly positive), and exciting (1 = not exciting at all, 7 = highly exciting) the imagined event was, and how expensive the imagined event would be (1 = not expensive at all (very cheap), 7 = highly expensive).

Consideration of Future Consequences (CFC) Scale. The CFC scale (Joireman, Shaffer, Balliet, & Strathman, 2012) consists of 14 statements that can be split into two subscales: CFC-immediate and CFC-future. An example of an item of the Immediate-subscale is "I only act to satisfy immediate concerns, figuring the future will take care of itself." An example of an item of the Future-subscale is "I consider how things might be in the future, and try to influence those things with my day to day behavior." Participants had to rate on a 7-point likert scale how characteristic each statement was of themselves (1 = extremely uncharacteristic of you, 7 = extremely characteristic of you). Items from the Immediate-subscale were reverse scored, and scores were averaged across all items for each participant. For a list of all 14 items, refer to Appendix A.

Post-task questions. Participants were asked: '*What did you think about when you answered the questions with cues (written in red) during the money task?*' and '*What did you think about when you answered the questions without cues during the money task?*'

Procedure

All participants completed all tasks in the following order: EFT event, cue and date generation, hypothetical money task, cue ratings, CFC scale, post-task questions. At the end of the session, participants were given a written and verbal debrief which explained the purpose of the current experiment. The entire session took approximately 45 minutes to complete.

Data analyses

Indifference points were plotted against their delays for each condition for each participant. Microsoft Excel Solver was used to fit the hyperbolic model (Equation 1.1) to the indifference points (expressed as a proportion of 'A') for each condition, and *k* and R^2 values were identified. We also used AUC as a second discounting rate measure (Equation 1.2).

One methodological issue in the current experiment is that the control delays were always slightly longer than the EFT delays for each participant. This could affect the trapezoid areas and the total possible area used to calculate AUC. The group median control delays were also longer than the group median EFT delays (except for the final delay) in Peters and Büchel's (2010) study. However, Peters and Büchel only used *k* as a discounting rate measure. Given the questionable validity of *k* in the EFT condition of the current experiment (see Results and Discussion below), we also wanted to be able to use AUC as a measure of discounting. We therefore calculated a new AUC value for the control condition for each individual that corrected for this confound. First, we calculated the indifference points predicted by the hyperbolic model using the delays from the EFT condition and the *k* value from the control condition. That is, we estimated indifference points at the delays used in the EFT condition if EFT cues were absent. These predicted indifference points were constrained to a maximum of 0.95, given that the observed indifference points in the EFT condition could not exceed 0.95 (the hyperbolic model can predict indifference points as high as 1). We then calculated the corrected-AUC values by using those indifference points predicted by the hyperbola as the y values and the EFT delays as the x values. Appendix B shows a worked example of how we calculated this corrected-AUC based on one participant's data.

Tests of normality (Shapiro Wilk and visual inspection of histograms depicting data distributions) showed that AUC values were normally distributed (*ps* > 0.05). The *k* values were not normally distributed (*ps* < 0.001). We therefore log transformed the *k* values which normalized the distribution (*p*s > 0.05). We used paired-samples t-tests to determine whether discounting rates were significantly different between the EFT and control conditions. Effect sizes were calculated using Hedges' *g* statistic for within subject analyses (see Chapter 3 [meta-analysis]).

In order to determine the size of the difference in discounting rate between the EFT and control delays, we calculated the log ratio of EFT AUC and control AUC for each participant (LOG [EFT AUC/control AUC]). We also calculated an imagery score for each participant by averaging cue ratings across events and the frequency and vividness dimensions (as done by Peters & Büchel, 2010). The log ratios were not normally distributed (*p* < 0.001). We therefore calculated Spearman's rank correlation coefficients between: 1) log ratios and CFC scores, and 2) log ratios and imagery scores.

We also calculated the size of the EFT effect at each delay range by calculating the log ratio of the EFT indifference point and the hyperbola-predicted indifference point in the control condition at the EFT delay. The hyperbola-predicted indifference points for the control condition were calculated as previously described, to correct for the fact that the control delays were always longer than the EFT delays. These log ratios were not normally distributed (*p*s > 0.05). We therefore used a Friedman test to determine if the delay range significantly affected the size of the EFT effect. Post hoc comparisons were made using Wilcoxon-signed rank tests, using a Bonferroni-corrected alpha level of 0.01 to correct for the number of comparisons.

Levels of unsystematic discounting and data exclusion

Unsystematic data are usually identified and excluded if: 1) Any indifference point was larger than the previous indifference point by more than 20% of the larger, delayed reward, and 2) The last indifference point was not less than the first indifference point by at least 10% the larger, delayed reward (Johnson & Bickel, 2008). It is common practice within the discounting literature to exclude participants with data meeting these criteria because it suggests that they did not understand the task or were not attentive to the task. A recent meta-analysis by Smith, Lawyer and Swift (2018) showed that the mean frequency of unsystematic delay discounting data was 18%. However, the percentage of participants with unsystematic discounting in the current experiment was notably higher (28%). Table 2.1 shows the percentage of participants with unsystematic discounting according to each criteria in each condition.

Table 2.1
Percentage of participants with unsystematic discounting (based on Johnson & Bickel's 2008 criteria) in the EFT and control conditions.

Condition	Criterion 1 only	Criterion 2 only	Both criteria
Control	0%	0%	3%
EFT	8%	15%	5%

As shown in Table 2.1, unsystematic discounting was mainly found in the EFT condition. Specifically, most participants failed criterion 2 only, and this was due to the fact that those participants continuously chose the larger, delayed reward, which produced a flat discounting function. An extremely effective EFT manipulation might produce no discounting at all, and therefore the higher percentage of unsystematic discounting may be explained as a consequence of the EFT manipulation, rather than participant misunderstanding or inattention. Stein et al. (2016) also reported little to no discounting in an EFT condition.

Unsystematic discounting as defined by criterion 1 in the EFT condition may also be a consequence of the EFT manipulation if the events were not matched on how positive, exciting or expensive they were for those participants. For example, one participant showed systematic discounting in the EFT condition until the final delay, where the indifference point greatly increased. The event at this delay was an overseas trip which was rated as more exciting and expensive than any of the other events. Previous studies assessing the effect of EFT on discounting rate have matched events on such dimensions, and have not reported high levels of unsystematic discounting. For example, Peters and Büchel (2010) had participants describe several events and corresponding cues for each delay, and only cues that were matched on such dimensions were subsequently used for the money discounting task. Peters and Büchel reported that only 8% of their participants had unsystematic data (based on both criteria overall).

We therefore only excluded participants if they showed unsystematic discounting in the control condition, given that this likely reflects participant misunderstanding or inattention. Only two participants met these exclusion criteria, and thus the final analysed sample size was 37.

We also formally tested whether the unsystematic discounting based on criterion 1 in the EFT condition was due to unmatched event ratings. First, we calculated the size of the difference between the observed EFT indifference point and the hyperbola-predicted EFT indifference point (log [EFT indifference point/hyperbola-predicted EFT indifference point]) for each delay for each individual. A positive log ratio indicates that the observed indifference point was greater than the hyperbola-predicted indifference point. We then calculated the positivity, expensiveness and imagery score ratings at the delay where the EFT indifference point positively deviated the most from the hyperbola-predicted indifference point(i.e., the greatest log ratio) and at the delay where the EFT indifference point deviated the least from the hyperbola-predicted indifference point (i.e., the smallest log ratio) for each individual. We then compared the event ratings for these two indifference points using Wilcoxon signed rank tests. If an event rating is significantly higher at the indifference points that positively deviated the most from the hyperbola compared to at the indifference points that deviated the least from the hyperbola, then this would indicate that participants had unsystematic discounting in the EFT condition because the events were unmatched on that particular attribute.

Results and Discussion

Discounting rates

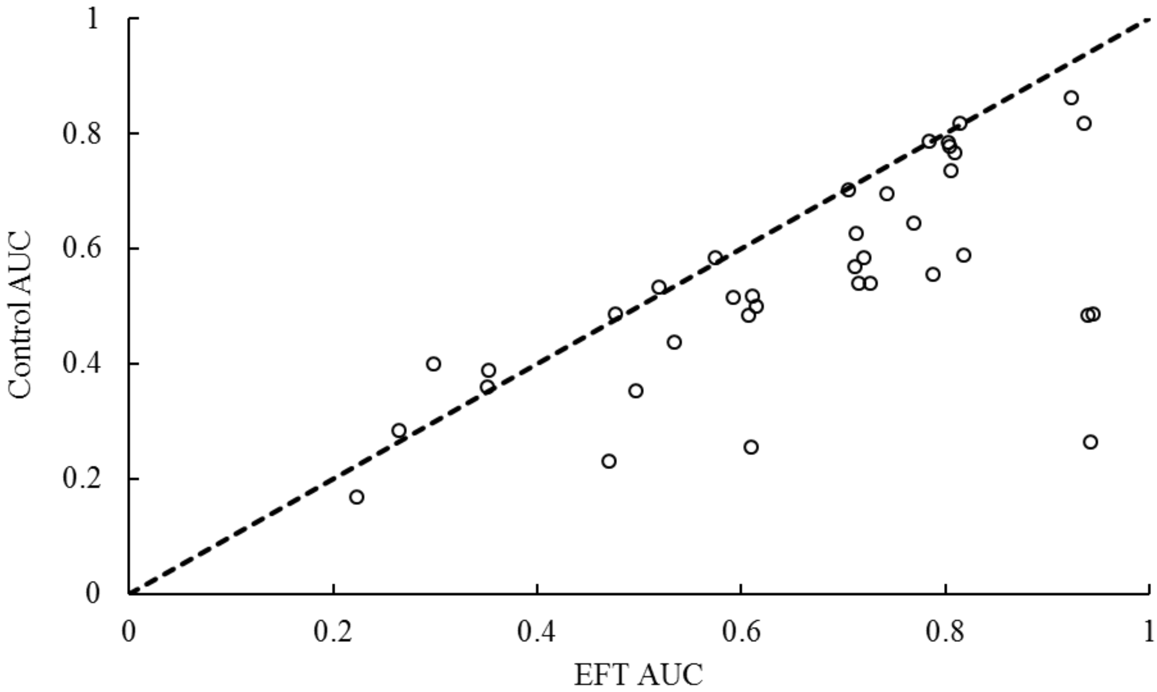

Figure 2.1. AUC values for the EFT delays plotted against AUC values for the control delays. Each circle represents one participant. Plotted circles below the dashed line represent participants with higher AUC values for EFT delays compared to control delays. Plotted circles above the dashed line represent participants with higher AUC values for control delays compared to EFT delays.

AUCs were significantly higher for the EFT delays (*M* = 0.66, *SD* = 0.20) compared to the control delays (*M* = 0.54, *SD* = 0.18), *t* (36) = 4.58, $p < 0.001$, and the size of this effect was moderate (*g* = 0.614). Figure 2.1 depicts the individual data where each circle represents one individual. The circles below the dashed line depict individuals who had higher AUCs for the EFT delays, whereas the circles above the dashed line depict individuals who had higher AUCs for the control delays. As shown in the figure, most participants (73%) had a higher AUC value for the EFT delays compared to the control delays. When we excluded participants who had unsystematic discounting in the EFT condition, the EFT effect remained significant but reduced in size (EFT *M* = 0.64, *SD* = 0.19; control *M* = 0.57, *SD* = 0.17; *t* (28) = 4.21, $p < .001$, *g* = 0.460).

Given that 15% of the participants showed no discounting in the EFT condition, we considered AUC to be a more valid discounting rate measure than *k*, given that it makes little sense to fit a non-linear model (from which *k* is revived from) to data that are in fact linear for a portion of the individuals. Additionally, the hyperbolic model poorly described discounting in the EFT condition for a further 21% of the individuals (R^2s < 0.5). This is not surprising given that most of these individuals had unsystematic discounting according to criterion 1, and the hyperbolic model assumes that indifference points decrease as the delay to the larger amount increases (with the rate of decrease being faster over initial increases in delay). Nevertheless, log *k* values produced results that were consistent with the AUC results, as log *k*s were significantly higher for the control delays (*M* = -2.24, *SD* = 0.43) than for the EFT delays (*M* = -2.59, *SD* = 55; *t* (36) = 4.34, $p < .001$, *g* = 0.682). Discounting in the control condition was well described by the hyperbolic model (*Mdn* $R^2 = 0.83$).

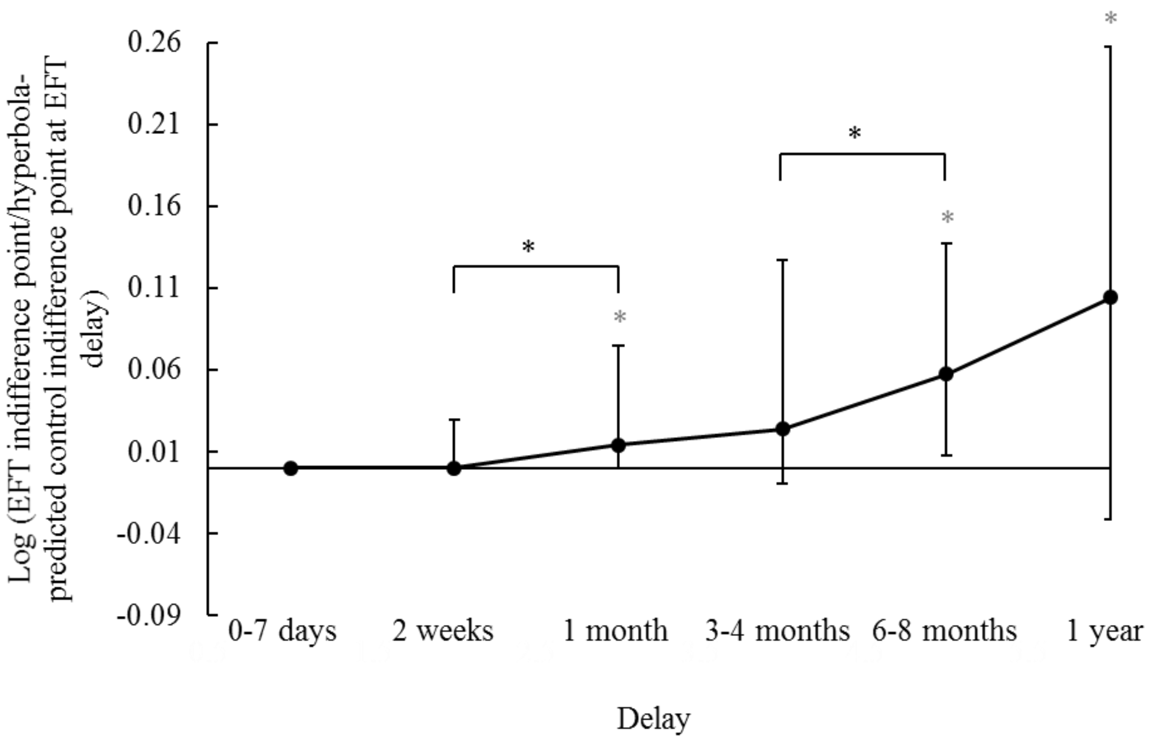

Figure 2.2. The group median size of the EFT effect (log [EFT indifference point/hyperbola-predicted control indifference point at the EFT delay]) as a function of delay. Points above the x-axis indicate that participants were more self-controlled for the EFT trials than for the control trials. The grey asterisks indicate that the size of the EFT effect was significantly different from zero, and the black asterisks indicate a significant difference between the two bracketed delays. The error bars represent the interquartile range.

Figure 2.2 shows the size of the EFT effect (log [EFT indifference point/hyperbola-predicted control indifference point at the EFT delay]) as a function of delay. A Friedman test showed a significant main effect of delay on the size of the EFT effect, *X* = 42.30, $p < 0.001$. As shown in the figure, EFT did not reduce discounting at the shorter delays (0-7 days and 2 weeks). One-sample Wilcoxon signed-rank tests showed that the size of the EFT effect was only significantly different from zero at the 1 month ($p = 0.004$), 6-8 month ($p = 0.001$) and

1-year ($p < 0.009$) delays. Wilcoxon signed rank tests also showed that the size of the EFT effect increased from the 2 week delay to the 1 month delay ($p < 0.001$) and from the 3-4 month delay to the 6-8 month delay ($p = 0.005$; all other *p*s between neighbouring pairs > 0.05).

Event ratings

There was no significant correlation between CFC scores and the size of the EFT effect according to both AUC (r_s (35) = -0.03, $p = 0.867$) and *k* (r_s (35) = 0.05, $p = 0.786$), in contrast to Benoit et al. (2011). There was also no significant correlation between imagery scores and the size of the EFT effect according to both AUC (r_s (35) = 0.26, $p = 0.262$) and *k* (r_s (35) = -0.28, $p = 0.093$). This finding is inconsistent with Peters and Büchel (2010), who found a significant positive correlation between the size of the EFT effect and imagery scores. Figure 2.3 shows scatter plots depicting the correlations.

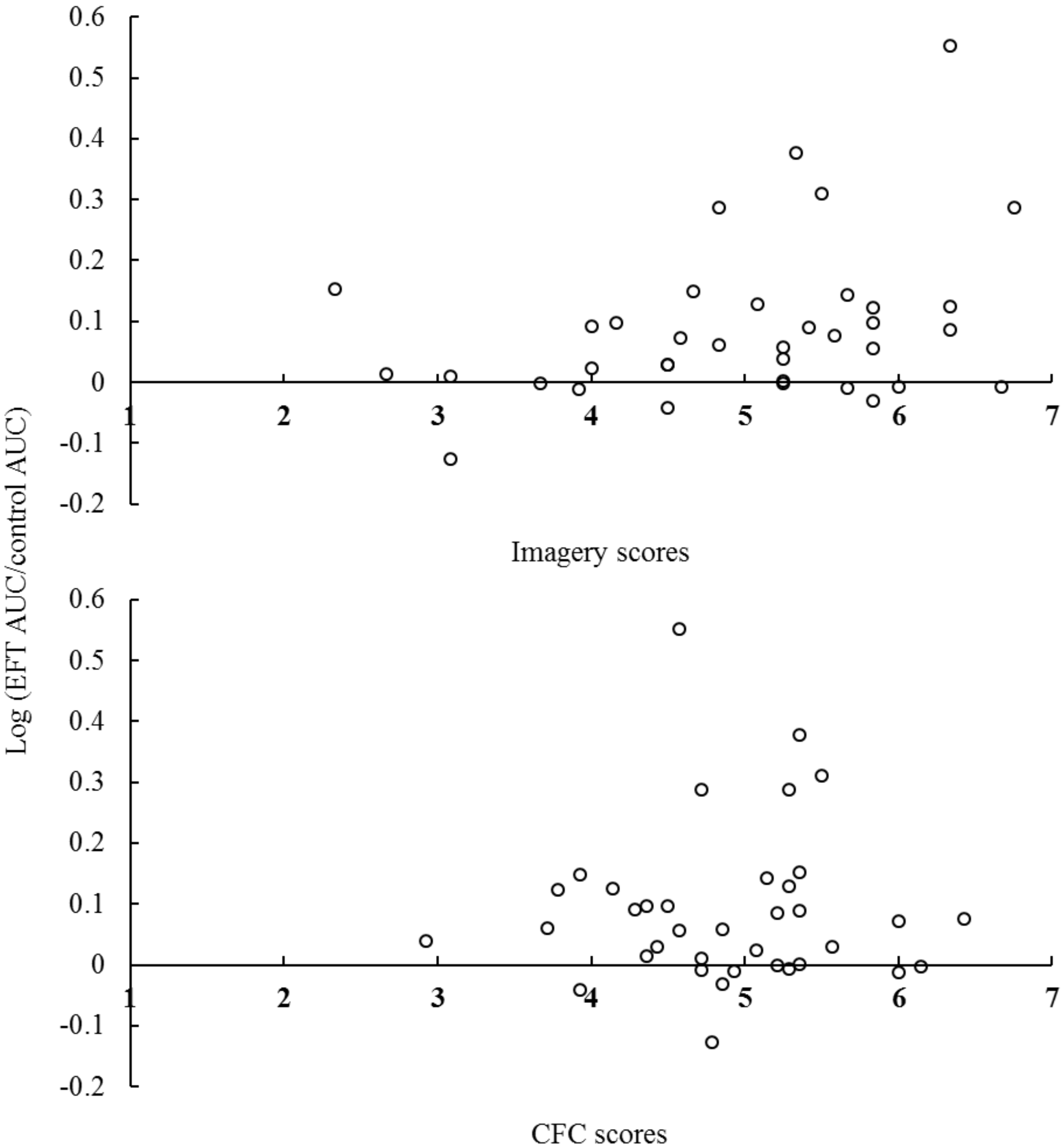

Figure 2.3. The size of the EFT effect (log [EFT AUC/control AUC]) as a function of imagery scores (top graph) and CFC scores (bottom graph). Circles above the x-axis are participants who were more self-controlled during the EFT condition than during the control condition.

As shown in the top graph of Figure 2.3, the non-significant correlation between the size of the EFT effect and imagery scores may be due to the restricted range of imagery scores. Few participants gave ratings of less than 4 for the imagery scale. In contrast, 37% of the participants in Peters and Büchel's (2010) study gave imagery scores below the midpoint of the scale, and these participants tended to have steeper discounting in the EFT condition (which would correspond to a negative log ratio), consistent with the distribution of the points evident in the top graph of Figure 2.3. As shown in the bottom graph of Figure 2.3, few participants also gave ratings of less than 4 for the CFC scale. However, this range of scores is consistent with the range of CFC scores in Benoit et al.'s (2011) study. Thus, our non-significant correlation between the size of the EFT effect and CFC scores cannot be attributed to a range problem. Rather, Benoit et al.'s significant correlation using a small sample size of 12 did not replicate using our larger sample size of 37[2].

There was no significant difference in any of the event ratings between the indifference points with the smallest EFT effect and the indifference points with the largest EFT effect (imagery ratings: largest effect *Mdn* = 5.5, smallest effect *Mdn* = 5, $p = 0.043$; expensiveness ratings: largest effect *Mdn* = 5, smallest effect *Mdn* = 4, $p = 0.078$; all other rating medians were identical). Although the difference for the imagery scores was significant at the 0.05 alpha level, this was not significant at the Bonferroni-corrected alpha level of 0.025 that was warranted to correct for the two tests.

Additionally, there was no significant difference in any of the event ratings between the EFT indifference points that positively deviated the most from the hyperbola and the EFT indifference points that deviated the least from the hyperbola (imagery ratings: largest effect *Mdn* = 5.5, smallest effect *Mdn* = 5, $p = 0.033$; expensiveness ratings: largest effect *Mdn* = 5, smallest effect *Mdn* = 4.5, $p = 0.334$; all other rating medians were identical). Although the difference for the imagery scores was again significant at the 0.05 alpha level, this was not significant at the Bonferroni-corrected alpha level of 0.025. These results do not support the conclusion that participants were unsystematic in the EFT condition according to criterion 1 due to the events being unmatched on a particular attribute. However, note that this could have only accounted for a small percentage of the participants, as only 13% of the entire sample were unsystematic according to criterion 1.

'What were you thinking about' questions

When asked about their decisions about EFT delays in the money task, most participants (71%) stated that they thought about how useful the amount of money would be for that particular event, even though they were told that they did not have to base their

---

[2] Benoit et al. (2011) used a 12-item version of the CFC scale (Strathman et al., 1994) that did not include the last two items of the 14-item version of the scale that we used (Joireman et al., 2012; see Appendix A). However, the correlation between the size of the EFT effect and CFC scores was still non-significant when we excluded those two items ($r_s(35) = -0.04$, $p = 0.837$).

decisions on the event. Only 8% of the participants said that they thought about possible events during the control delays. The rest of the participants stated that they thought about whether the wait was worth the larger amount of money. These findings indicate that EFT generally only occurred during the EFT trials even though EFT trials and control trials occurred in close proximity to each other.

Conclusion

The aim of Experiment 1 was to test whether we could replicate the effect of EFT on delay discounting. Discounting rates were significantly lower in the EFT condition than in the control condition, consistent with previous studies. In Chapter 3, we further investigated the reliability of the EFT effect by conducting a meta-analysis of 40 independent studies.

Chapter 3: Meta-analysis

Rung and Madden (2018b) recently published a review and meta-analysis assessing various experimental reductions of delay discounting rate, including EFT. They found a significant, small effect of EFT on delay discounting rate (Hedges' g = 0.38), with little variability between studies. However, Rung and Madden's meta-analysis only included ten EFT studies, given that their search concluded in 2016. The effect of EFT has since become a research area of major interest, and many more studies have since been conducted. Chapter 3 therefore provides a more up-to-date meta-analysis on the effect of EFT on delay discounting rate, with four times the number of independent effect sizes examined. We were also able to assess potential moderators of the effect sizes, given the wider range of studies included. Below we outline moderators of interest.

Study design

In within-subjects designs each subject serves as their own control. Therefore, the standard deviation of the difference in means (i.e., the denominator of the effect size equation) is typically smaller for within-subjects designs, which will produce a larger effect size compared to between-subjects designs (Lakens, 2013). For this reason, when conducting a meta-analysis including both of these types of designs, Morris and deShon (2002) recommend that: 1) effect sizes calculated from within-subjects and between-subjects designs are transformed into a common metric, and 2) experimental design is examined as a moderator of the effect sizes. If significant differences in effect sizes are found between the two designs, then meta-regression may be used where design and all other significant moderators are analyzed simultaneously while controlling for each other. Thus, we examined experimental design as a moderator in the current meta-analysis.

Design might also be a significant moderator of the effect sizes due to features of the different designs themselves, rather than for statistical reasons. For example, effect sizes might be larger in within-subjects designs because the researcher's hypothesis is more apparent when participants experience both conditions, and thus demand characteristics could have a greater effect on the results (Sawyer, 1975). As noted in the general introduction, demand characteristics, where participants change their behavior to conform to the researcher's expectations, are a concern within EFT procedures. If a within-subjects design increases the likelihood of demand characteristics, then this could, in turn, inflate the EFT effect.

Alternatively, the EFT effect might be smaller in within-subjects design studies due to order effects. An order effect is when the effect of an experimental manipulation is influenced by the order in which participants experience the conditions (Charness, Gneezy, & Kuhn, 2012). Such order effects have been found in delay discounting research (e.g., Dai, Grace, & Kemp 2009; Hinvest & Anderson, 2010; Olsen et al., 2018; Yi et al., 2017). For example, Yi et al. (2017) found that EFT (and other similar experimental manipulations) affected discounting for participants who experienced the control condition first, but not for participants who experienced the control condition second. This additional source of variability within a study could reduce the overall size of an effect.

Discounting rate measure

As previously noted in the general introduction, the effects of experimental manipulations on delay discounting rate can differ depending on whether *k* or AUC was used (Olsen et al., 2018; Smith & Hantula, 2008). For example, Olsen et al. (2018) found that reinforcer magnitude affected delay discounting of academic outcomes according to *k*, but not according to AUC. There are also additional, atheoretical ways of measuring discounting, and it is not clear how these measures might impact the magnitude of an EFT effect.

Control condition

As described in the general introduction, EFT studies have used a wide range of control conditions, and these different control conditions provide information about the importance of the episodic thinking and future thinking components of EFT. Findings that EFT reduces delay discounting compared to semantic future thinking (e.g., Benoit et al., 2011) show that episodic thinking is more effective at reducing delay discounting rate than semantic thinking. Additionally, findings that EFT reduces delay discounting compared to EPT (e.g., Daniel, Sawyer, Dong, Bickel, & Epstein, 2016) show that thinking about the future is more effective at reducing delay discounting than thinking about the past. We therefore assessed control type as a moderator of the effect sizes in order to better understand the relative importance of the components of EFT. For example, if the current meta-analysis shows that the EFT effect is larger when EPT (rather than semantic future thinking) control conditions are used, then this would suggest that thinking about the future (rather than episodic thinking) is the *more important* component of the EFT effect.

Additional moderators of interest

Additional moderators of interest were variables that are known to influence the subjective amount of the larger, delayed reward (i.e., delay length) or delay discounting rate (i.e., objective amount of the larger, delayed reward, age of the participants, reward type, characteristics of participants). In cases where subjective amount is smaller (or delay discounting rate is higher), EFT might have a larger effect because it is able to produce a greater increase in subjective value (or greater decrease in discounting rate) compared to baseline. One such variable is delay-length; longer delays might predict a larger EFT effect because subjective value tends to be lower at longer delays. Previous research has shown that the effect of EFT on delay discounting rate may depend on the delays used (Snider et al., 2016; Stein et al., 2016). For example, Snider et al. (2016) found that indifference points were significantly higher in an EFT condition compared to control, but only at the longest delays used (at 1 month, 3 months and 1 year, but not at 1 day or 1 week). Experiment 1 of the current thesis also found that EFT only reduced subjective value at these longer delays.

Delay discounting rate decreases with increasing delayed amounts (*the magnitude effect*; e.g., Benzion, Rapoport, & Yagil, 1989; Green, Myerson, & McFadden, 1997; Johnson & Bickel, 2002) and with increasing age of the participants (Green, Fry, & Myerson, 1994; Green et al., 1999). We therefore predicted that larger, delayed amounts and older age would be associated with smaller EFT effects, given that smaller changes in discounting rate relative to baseline are possible. Additionally, EFT might be less effective in older adults due to reduced executive functioning (Buckner, 2004; Sasse, Peters, & Brassen, 2017) and reduced ability to generate detailed future events (Addis, Wong, & Schacter, 2008), compared to younger adults. Interestingly, Sasse et al. (2017) found no significant effect of EFT on delay discounting rate in healthy, older individuals (mean age = 66.5).

Fourthly, given that discounting rate is known to be steeper in individuals with particular characteristics (e.g., substance-use or dependence [Madden et al., 1997; Mitchell, 1999; Petry, 2001], obesity [Davis et al., 2010; Weller et al., 2008] and problem gambling disorder [Dixon et al., 2003]) compared to healthy controls, we examined participant characteristics as a moderator of the EFT effect. As with other variables that produce steeper discounting, larger EFT effects might be possible for such participants. Alternatively, some subsets of individuals may show smaller EFT effects. For example, substance dependence is associated with reduced working memory capacity (Bechara & Martin, 2004), which is related to reduced EFT effectiveness (Lin & Epstein, 2014). Indeed, Snider at al. (2016) found that higher discounting rates in an EFT condition was associated with more severe alcohol abuse among individuals with alcohol-dependence. Additionally, higher body mass index is associated with decreased cerebral blood flow in particular brain regions that have been implicated in the EFT effect (Peters & Büchel, 2010; Willeumier, Taylor, & Amen, 2011), which may reduce EFT effectiveness in obese/overweight individuals. Although Daniel et al. (2013a) found no difference in EFT effectiveness between obese/overweight and lean individuals, they also found no overall difference in discounting rate between the two groups, in contrast to other studies (Davis et al., 2010; Weller et al., 2008).

Fifthly, we examined reward type as a moderator of the EFT effect. The types of rewards used in delay discounting tasks can be categorized as hypothetical, potentially real, or real (Reynolds, 2006). Most researchers use hypothetical reward tasks, where participants do not receive any money, but are instructed to make their choices as if they would receive the outcome they selected. In potentially-real reward tasks, participants experience only one or some of the outcomes they select, selected at random by the experimenter. In real reward tasks, participants experience all outcomes they select. Research shows that hypothetical rewards are discounted similarly to both real (Lagorio & Madden, 2005) and potentially-real (Johnson & Bickel, 2002; Madden et al., 2003) rewards. Thus, EFT effects may be comparable among the three reward types, but this has not previously been confirmed.

EFT valence as a moderator

To reiterate, both Liu et al. (2013) and Zhang et al. (2018) found that while thinking about positive future events reduced delay discounting, thinking about negative future events increased delay discounting, and thinking about neutral future events had no effect. Therefore, we intended to examine EFT valence as a moderator of the effect sizes. However, this analysis was not possible given that few studies assessed the effects of negative or neutral events (See Method [Overview of Analyses]).

Summary of aims

In summary, we investigated the average size of the effect of EFT on delay discounting rate. We also aimed to assess the robustness of the EFT effect across various methodological features and participant characteristics. These aims were achieved by conducting moderator analyses. Moderators examined were: 1) study design, 2) discounting rate measure type, 3) control type, 4) longest delay to the larger, delayed reward, 5) the amount of the larger, delayed reward, 6) average age of the participants, 7) participant characteristics, and 8) reward type.

Method

Selection of studies

To be included, studies had to compare the effects of EFT and a control on delay discounting rate, using an experimental design. EFT was defined as a manipulation where participants were required to imagine (or list) personal, future events prior to completing a delay discounting task. Instructions given to participants had to ask them to engage in episodic thinking (i.e., thinking or writing about a personal event, localized in time with a scene unfolding) with a future component. The control condition was defined as a condition that was not an EFT manipulation. Delay discounting rates needed to be obtained from a task in which participants made choices between smaller, sooner rewards and larger, delayed rewards.

A concern in meta-analyses is that effect sizes may be biased towards statistically significant results because studies with statistically non-significant results are less likely to be published ("the file drawer problem"; Rosenthal, 1979). To reduce the file drawer problem, we included both published and unpublished studies in our meta-analysis. Published and unpublished studies were obtained by searching the following databases: *Google Scholar*, *PsycInfo*, *ProQuest Research Library* and *PsycARTICLES*, and also from reference lists of obtained studies. Searches were conducted using the following key terms: *("delay discounting" OR "temporal discounting" OR "impulsive choice" OR "delayed gratification" OR intertemporal OR "impulsive decision making") AND ("episodic future thinking" OR "episodic prospection" OR "mental time travel").* We obtained further unpublished studies by emailing relevant authors. The literature search concluded in October 2018, and was limited to studies written in English.

In addition to failures to meet the inclusion criteria, studies were excluded if they did not report sufficient quantitative data to estimate effect sizes (and the authors could not be contacted for this information). Figure 3.1 is an adapted version of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA; Moher, Liberati, Tetzlaff, & Altman, 2010). The PRISMA is a flow diagram which depicts the study selection process. In total, the final sample included effect sizes from 40 experiments (k = 40) obtained from 36 studies.

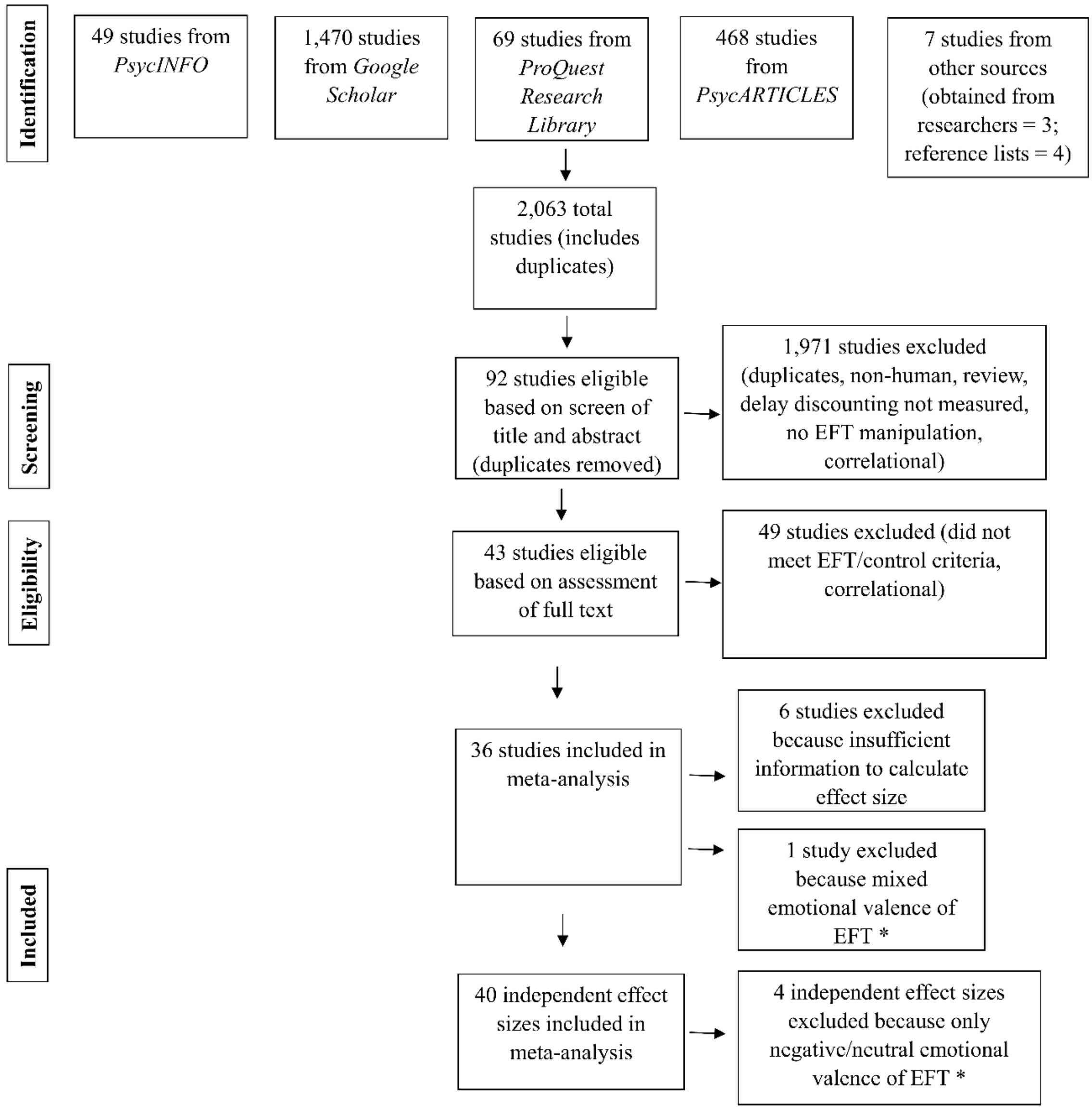

Figure 3.1. Adapted PRISMA flow diagram. * indicates that these effect sizes were initially coded and are included in interrater reliability analyses, but excluded in the meta-analysis (See Method [Overview of Analysis] section).

Effect size and variance calculations

We calculated effect sizes (Hedges' *g* [converted from Cohen's *d*]) of the difference in discounting rate between EFT and control conditions. Cohen's *d*, Hedges' *g* and the sampling variance of Hedges' *g* were calculated differently depending on the study design. When data were not provided in the manuscript and the author could not be contacted, we estimated means and standard deviations from published figures using WebPlotDigitizer (Rohatgi, 2018).

Between-subjects designs. When the means and standard deviations were available, Equation 3.1 was used to calculate Cohen's *d* for a between-subjects design (d_B) (Lakens, 2013; Morris & deShon, 2002):

$$d_B = \frac{M_1 - M_2}{\sqrt{\frac{(n_1-1)SD_1{}^2 + (n_2-1)SD_2{}^2}{n_1+n_2-2}}} \qquad \text{[Equation 3.1]}$$

Where M_1 and M_2 are the means of each group, n_1 and n_2 are the number of participants in each group, and SD_1 and SD_2 are the standard deviations of the mean for each of those groups.

If a study using a between-subjects design only reported the results of a *t*-test, then Equation 3.2 was used to calculate Cohen's *d* (Lakens, 2013; Morris & deShon, 2002):

$$d_B = t \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad \text{[Equation 3.2]}$$

Where t is the t test statistic.

Cohen's *d* tends to overestimate effect sizes when sample sizes are small (Cumming, 2011). To correct for this sample size bias, all Cohen's d_B values were converted into Hedges' g (g_B) using Equation 3.3 (Cumming, 2011; Lakens, 2013):

$$g_B = d_B \times \left(1 - \frac{3}{4(n_1+n_2-2)-1}\right) \qquad \text{[Equation 3.3]}$$

Where d_B is Cohen's *d* calculated from either Equation 3.1 or Equation 3.2.

The sampling variance of between-subjects Hedges' g (Var_{g_B}) was calculated using Equation 3.4 (Morris & deShon, 2002):

$$Var_{g_B} = \left(\left(\frac{1}{\left(\frac{(n1\times n2)}{(n1+n2)}\right)}\right) \times \left(\frac{((n1+n2)-2)}{((n1+n2)-4)}\right) \times \left(1 + \left(\frac{(n1\times n2)}{(n1+n2)}\right) \times (g_B^2)\right)\right) - \left(\frac{(g_B^2)}{\left(\left(1 - \frac{3}{(4\times((n1+n2)-2)-1)}\right)^2\right)}\right)$$

[Equation 3.4]

Where g_B is Hedges' g calculated from Equation 3.3.

Within-subjects designs. When the means and standard deviations were available, Equation 3.5 was used to calculate Cohen's *d* for a within-subjects design (d_W) (Lakens, 2013; Morris & deShon, 2002):

$$d_W = \frac{M_1 - M_2}{\sqrt{SD_1^2 + SD_2^2 - 2 \times r \times SD_1 \times SD_2}} \qquad \text{[Equation 3.5]}$$

Where M_1 and M_2 are the means of each condition, SD_1 and SD_2 are the standard deviations of the mean for each of those conditions, and *r* is the coefficient of the correlation between those two conditions.

If a study using a within-subjects design only reported the results of a *t*- test, then Equation 3.6 was used to calculate Cohen's *d* (Lakens, 2013; Morris & deShon, 2002):

$$d_W = \frac{\mathrm{t}}{\sqrt{N}} \qquad \text{[Equation 3.6]}$$

Where *t* is the t test statistic and *N* is the number of participants.

Two within-subjects design studies reported non-normal data that could not be normalized using log-transformation methods (Olsen 2018; Peters & Büchel, 2010 Experiment 2[3]). In these cases, the effect size *r* (r_{ES}) was calculated from a Wilcoxon signed rank test using Equation 3.7 (Fritz, Morris & Richler, 2012):

$$r_{ES} = \frac{Z}{\sqrt{\mathrm{N}}} \qquad \text{[Equation 3.7]}$$

Where *Z* is the *Z* score from the Wilcoxon signed rank test and *N* is the number of observations (i.e., the number of participants multiplied by 2). r_{ES} was then converted to the Cohen's d_W metric using Equation 3.8 (Fritz et al., 2012):

$$d_W = \frac{2r_{ES}}{\sqrt{1 - r_{ES}^2}} \qquad \text{[Equation 3.8]}$$

For within-subjects designs, all Cohen's *d* values were converted to a between-subjects design metric (d_{WC}) using Equation 3.9 (Lakens, 2013; Morris & deShon, 2002):

$$d_{WC} = d_W \times \sqrt{2(1 - r)} \qquad \text{[Equation 3.9]}$$

Where d_W is Cohen's *d* calculated from either Equations 3.5, 3.6, or 3.8, and *r* is the coefficient of the correlation between the EFT and control conditions. We chose Equation 3.9 over other calculations (see Lakens, 2013 for discussion) because this allowed us to calculate

---

[3] These data were obtained from the authors where *k* values were fitted differently to the *k* values reported in the original paper (see Peters, Miedl, & Büchel, 2012).

Cohen's *d* from within-subject design studies where the means and standard deviations were either inappropriate descriptions of the distribution (i.e., non-normal data) or unknown (i.e., only a *t*- test was reported).

When *r* was unavailable, the corresponding author was contacted. If the author did not respond, then *r* was estimated using Equations 3.10 and 3.11 (Morris & deShon, 2002):

$$r = \frac{SD_1^2 + SD_2^2 - SD_D^2}{(2)(SD_1)(SD_2)} \qquad \text{[Equation 3.10]}$$

$$SD_D^2 = \frac{N(M_1 - M_2)^2}{t^2} \qquad \text{[Equation 3.11]}$$

Where *N* is the number of participants and SD_D^2 is the variance of difference scores.

Finally, all Cohen's *d* values for within-subjects designs were converted into Hedges' *g* (g_W) using Equation 3.12 (Cumming, 2011; Lakens, 2013):

$$g_W = d_{WC} \times \left(1 - \frac{3}{4(N-1)-1}\right) \qquad \text{[Equation 3.12]}$$

Where d_{WC} is Cohen's *d* calculated from Equation 3.9, and *N* is the number of participants.

The sampling variance of within-subjects Hedges' *g* (Var_{g_W}) was calculated using Equation 3.13 (Morris & deShon, 2002):

$$Var_{g_W} = \left(\left(\frac{1}{N}\right) \times \left(\frac{(N-1)}{(N-3)}\right) \times (1 + N \times (g_W^2))\right) - \left(\frac{(g_W^2)}{\left(\left(1 - \frac{3}{(4 \times (N-1) - 1)}\right)^2\right)}\right)$$

[Equation 3.13]

Where g_W is Hedges' *g* calculated from Equation 3.12, and *N* is the number of participants.

Coding procedure

For the 38 accepted studies[4], effect sizes, effect size variance and study descriptors were coded using a standardized coding manual (see Appendix C). If EFT reduced delay discounting (i.e., produced a higher AUC/higher indifference points/higher percent of larger, delayed rewards chosen/a higher reward index/a lower k) compared to the control condition, effect sizes were coded as a positive value. If EFT increased delay discounting effect sizes were coded as a negative value. Two sample characteristics and eight methodological characteristics were coded for each effect size.

Sample characteristics. Participant characteristics were coded as either a general population (i.e., no specific type of participant recruited), alcohol-dependent, nicotine-dependent, other substance use disorder, amnesia, or obesity/overweight. If the coder did not determine that the participants fit into any of these six categories, then they coded them as 'other' and provided a written description (this rule also applied for 'other' categories of the other study descriptors below).

Coders also recorded the average age of the participants. If these were reported separately within the study for EFT and control groups, then coders calculated the weighted average age of the groups.

Methodological characteristics. Reward type was coded as either real, potentially real or hypothetical (see Introduction). None of the rewards were coded as 'real', leaving only two categories (potentially real and hypothetical).

Control type was coded as either 'no thinking', episodic past thinking, episodic present/recent thinking, semantic/non-episodic, events from a story, or 'other.' 'No thinking' (NT) control conditions were defined as those that included no extra task or prompt during the delay discounting task (such as the standard delay discounting task, as in Peters & Büchel, 2010), or a task before the delay discounting task that did not target components relevant to EFT. For example, listening to music (Lempert, Porcelli, Delgado, & Tricomi, 2012) or completing a personality questionnaire (Wu, Cheng, & Chiou, 2017). Episodic

[4] Cheng, Shein, and Chiou (2012) is included as one of these 38 studies. However, this study was later removed (and therefore is one of the six studies excluded from the meta-analysis as indicated on the PRISMA flow diagram in Figure 1) because we had reason to believe that the available statistics within the paper violated assumptions of normality (see Rung & Madden 2018b, Footnote 4). Parthasarathi et al. (2017) was also included as one of these 38 studies, however, this study was later removed from the meta-analysis given that their EFT manipulation likely required a mix of positively- and negatively-valenced EFT (see Figure 3.1 and Overview of Analysis section).

present/recent thinking was defined as thinking about events "within the next (present) or past (recent) 24 hours." Past events exceeding 24 hours were coded as episodic past thinking (EPT), rather than episodic recent thinking. Semantic/non episodic control conditions were defined as those in which participants thought about generalized *information* (which may or may not be localized in time, e.g., "estimate what you could buy at the pub with £35 in 180 days"; Benoit et al., 2011) or a generalized *activity* not localized in time. An example of the latter is asking participants to imagine an event as it would typically unfold, rather than to remember a particular occasion when they did this activity (e.g., using a pencil/opening a cabinet; Bulley et al., 2019). Bulley et al. (2019) called this a "neutral, non-temporal" condition, which we coded as semantic/non-episodic. These two examples (Benoit et al., 2011; Bulley et al., 2019) differ from episodic thinking where participants think about a personal *event/activity*, localized in time with a scene unfolding. An 'events from a story' control condition was defined as events that participants read about from someone else's life (e.g., from an online travel blog or a book).

The longest delay used in the delay discounting task was recorded in days. The delayed amount used in the delay discounting task was originally recorded in the currency reported in the study, however these were later converted to a common metric in two different ways. First, we converted all currencies to US$, given that the majority were already in this currency. However, a disadvantage of this strategy is that it does not take into account the relative cost of living in that particular country. Thus, we also calculated the original amounts as a proportion of the country's median house-hold income. For comprehensiveness, results of delayed amount as a moderator were assessed both ways (see Results). For studies where more than one delayed amount was used in the delay discounting task, we recorded the largest delayed amount.

Study design was coded as either within-subjects or between-subjects. Delay discounting rate measure was coded as either *k,* AUC, or 'other.' Coders specified the EFT valence (i.e., whether participants were instructed to think about positive, neutral or negative events) on the form on which they recorded effect size information. Finally, each study was coded as either published or unpublished. Unpublished studies here refer to studies that were not published in a journal article (i.e., theses, manuscripts currently under review for publication, or conference posters) at the time of the literature search.

Interrater reliability (IRR)

I coded all 38 studies, and the first 27 studies were also coded by a research assistant. I developed, tested and refined the coding manual and trained the research assistant on its use. The remaining 11 studies were also coded by a second research assistant. I trained the second research assistant using a refined version of the standardized coding manual developed by me and the first research assistant (see Appendix D for further information). Comparisons between my ratings and the two research assistants' ratings are reported below. IRR analyses (Absolute Agreement Intra-class correlations [ICCs] using a One-Way Random model for continuous variables and Cohen's Kappa for categorical variables) are reported for all but five of the 38 studies that were used for training purposes.

The three raters coded 66 common effect sizes, with excellent levels of agreement (ICC based on a single rater = 0.982; Koo & Li, 2016). When coding, raters must identify which analyses in a paper report appropriate effect sizes that can be coded. In addition to the common effect sizes, 14 additional effect sizes were identified and coded by one rater but the other rater overlooked them or chose not to include them. A consensus was reached regarding the appropriateness of these additional effect sizes, and these effect sizes were re-calculated by both raters (although not independently) to ensure accuracy. For the other continuous variables coded, ICC values ranged from 0.976 to 1 (median ICC = 1), indicating excellent reliability. For categorical variables, Cohen's Kappa ranged from 0.825 to 1 (median ICC = 1), indicating almost perfect reliability (Landis & Koch, 1977). A consensus rating was reached for any disagreements for all variables. There were five studies that the raters could not code an effect size for due to insufficient quantitative data (De La Vega, 2011; Kwan et al., 2015; Sasse et al., 2017; Sasse, Peters, Büchel, & Brassen, 2015; Stein, Daniel, Epstein, & Bickel, 2015).

The IRR analyses of all of the effect sizes and their variance are based on Cohen's *d* calculations. I later transformed these into the Hedges' *g* metric by using Equations 3.3 and 3.12. Given that these transformations could be achieved in Microsoft Excel® using formulae based on the already agreed upon data, IRR analysis of the final Hedges' *g* values was unnecessary.

Overview of analyses

We used a random-effects model for the overall analysis (as recommended by Borenstein, Hedges, Higgins, & Rothstein, 2009 and Borenstein & Higgins, 2013), given that we expected true between-study variation in effect sizes. Each effect size was weighted by the inverse of its variance, and 95% confidence intervals were calculated for the weighted average effect size. All meta-analytic procedures were conducted using Comprehensive Meta-Analysis software (CMA version 3; Borenstein, Hedges, Higgins, & Rothstein, 2013).

Multiple effect sizes. We were able to code multiple, statistically dependent effect sizes for the majority of studies. This occurred most frequently because a study compared multiple control conditions to the same EFT condition, or multiple discounting rate measures were used within a study. As these effect sizes were from the same sample, they are statistically dependent, which violates the assumption of independence in meta-analyses, and may produce bias by assigning more weight to studies with more effect sizes (Lipsey & Wilson, 2001). For the overall analysis, we addressed this problem by averaging multiple, dependent effect sizes within a study. When combining variance calculations, the correlation among the different variables should be taken into account (Borenstein et al., 2009). Given that we did not know the correlation among variables, we simply averaged the variance calculations, which assumes a correlation coefficient of 1 and could inflate type 2 error. However, this is considered to be the more conservative approach (Borenstein et al., 2009).

Moderator analyses were conducted in one of two ways when there were multiple dependent effect sizes. If the value of the moderator was consistent across all dependent effect sizes within a study (e.g., design), then we used the averaged effect sizes as described above. However, if the value of a moderator was inconsistent across dependent effect sizes within a study, then we only allowed one of those effect sizes to contribute to the moderator analysis, as suggested by Lipsey and Wilson (2001). This effect size was selected based on the distribution of that moderator across studies. For example, if a study reported dependent effect sizes for two types of control conditions, then the control type that was represented by a smaller number of studies was selected, with the aim of producing a more even distribution of control type across studies.[5]

[5] Note that we also considered Cooper’s (1998) shifting unit of analysis technique, and this did not change results or improve the distribution of moderators across studies. This technique is where multiple, dependent effect sizes within a study are allowed to contribute to a particular moderator analysis, as long as those effect sizes represent different categories of that moderator. For example, if a study contained dependent effect sizes for two different types of control conditions, then those effect sizes would not be averaged together during the moderator analysis assessing control type. However, for all other moderator analyses, those effect sizes would be averaged so that that study only contributed one effect size.

We entered independent subgroups within the same study (e.g., gamblers and non-gamblers, as in Wiehler, Petzschner, Stephan, & Peters, 2017) as subgroups and used the study as the unit of the overall analysis by merging the effect sizes from the subgroups (as recommended by Borenstein et al., 2009). However, we used subgroup as the unit of analysis when assessing participant characteristics as a moderator. That is, we treated the subgroups as separate studies, but only for studies that had different participant characteristics across subgroups.

Separate experiments with separate groups of participants within a study were entered as separate studies.

Moderator analyses. For categorical moderators (i.e., subgroup analyses), we used a meta-analytic analogue of an ANOVA, which partitions the total variance into the variance between subgroups/categories of a moderator (i.e., the variance explained by the moderator) and the leftover variance that exists within subgroups/categories of a moderator. The significance of these two components is then tested using a chi square distribution (the Q test), with degrees of freedom equal to the number of subgroups of the moderator minus 1. A statistically significant between-level Q indicates that the moderator explains a significant proportion of the variability across studies. We used a mixed effects model (random-effects within subgroups and fixed effects across subgroups) as recommended by Borenstein et al. (2009) and Borenstein and Higgins (2013). We pooled the within-group estimates of the between-study variance (τ^2) across the subgroups of studies so that the categorical moderator analyses would be comparable to the multiple meta-regression analysis (see below). This approach also provides better τ^2 accuracy when there are five or fewer studies per subgroup (Borenstein et al., 2009).

For continuous moderators, we used meta-analytic regression. Meta-regression is comparable to regression analysis, where the outcome is the average effect size and moderators are predictors (covariates). Similar to regression, meta-regression computes a beta coefficient and an R^2, which is the proportion of between-study variance explained by the covariates. We used a method of moments (MM) to estimate τ^2 because it does not make any assumptions about the distribution of the effect sizes, as opposed to other methods which assume a normal distribution.

To test the significance of moderators simultaneously (while holding other moderators constant), we used multiple-meta-regression (MM random effects). For each categorical covariate in the regression, we created dummy variables where one category served as the reference group for the other categories. We also assessed whether any moderators included in the regression model were highly correlated, given that this can create statistical confounds among moderators (Lipsey, 2003).

Exclusion of effect sizes due to poorly distributed variables across studies. Whether participants were asked to think about positive, negative, or neutral future events (EFT valence) was initially a moderator of interest and was therefore coded for each study. However, given that very few studies assessed the effects of only negative (k = 3) or only neutral (k = 2) future events, we could not formally assess EFT valence as a moderator. This is because when a moderator is unevenly distributed across studies (i.e., when many more studies are in one category than others, or one category has very few cases), it is difficult to determine whether the results of that moderator analysis are meaningful or spurious (Lipsey, 2003). We therefore excluded effect sizes from non-positively valenced EFT from our overall meta-analysis, given that we were unable to statistically control for this additional source of variability. We excluded one additional study because raters agreed that the EFT manipulation asked participants to consider events that were likely somewhat positive and somewhat negative (Parthasarathi, McConnell, Luery, & Kable, 2017).

Secondly, we excluded an effect size from one study with participants with amnesia (Palombo et al., 2015)[6], as this condition produces severely impaired episodic prospection (Klein, Loftus, & Kihlstrom, 2002; Kwan, Carson, Addis, & Rosenbaum, 2010). We also excluded effect sizes from one study with participants with Subjective Cognitive Decline (SCD; Hu et al., 2017b). There was not a sufficient number of studies available to directly assess the impact of these potential or actual cognitive impairments on the effect of EFT on delay discounting.

The coded effect sizes and study characteristics of effect sizes excluded from the meta-analysis are shown in Table 3.1.

[6] Kwan et al. (2015) also used amnesiac participants, however we were unable to code an effect size from this study.

Table 3.1.

Coded effect sizes and study characteristics of effect sizes excluded from the meta-analysis based on EFT valence and participant characteristics.

Study name	Published	Design	DRM	Control	Participant type	Average age (years)	Reward	Delayed amount (converted to US$)	Longest delay (days)	Hedges' *g*	Reason for exclusion
Bulley et al. (2019)	No	BS	*k*	ST	GP	19.72	H	AU$85 (60.84)	186	0.519*	Negative valence
Hu et al. (2017b)	Yes	WS	AUC & k	NT	SCD	68.29	H	€200 (228.56)	365	0.010#	SCD
Lempert et al. (2012)	Yes	BS	%SSR	Other	GP	20.46	PR	US$10	365	-1.594*	Negative valence
Liu et al. (2013.2)	Yes	WS	%SSR	NT	GP	20.74	PR	CNY67.50 (10.01)	30	-0.506	Negative valence
Liu et al. (2013.3)	Yes	WS	%SSR	NT	GP	21.48	PR	CNY67.50 (10.01)	30	0.012	Neutral valence

Study name	Published	Design	DRM	Control	Participant type	Average age (years)	Reward	Delayed amount (converted to US$)	Longest delay (days)	Hedges' *g*	Reason for exclusion
Palombo et al. (2015)	Yes	BS	RRI	NT	Amnesia	61.22	H	US$58	730	-0.013#	Amnesia
Parthasarathi et al. (2017)	Yes	BS-PP	*k*	NT	GP	24.70	PR	US$35	180	-0.035	Mixed valence
Zhang et al. (2018.2)	Yes	WS	%SSR	NT	GP	21.32	PR	CNY172 (25.51)	120	-0.695	Negative valence
Zhang et al. (2018.3)	Yes	WS	%SSR	NT	GP	21.19	PR	CNY172 (25.51)	120	0.018	Neutral valence

Note. WS = within-subjects; BS = between-subjects; BS-PP = between-subjects pre-test post-test (see Morris & deShon, 2002 and the coding manual in Appendix C for the Cohen's *d* calculation used for this particular design).
DRM = discounting rate measure; %SSR = percent (or proportion) smaller, sooner rewards chosen; RRI = re-calibrated reward index (difference between actual accumulated reward and minimum accumulated reward possible divided by the difference between the maximum possible and minimum possible accumulated rewards).
ST = semantic thinking; NT = 'no thinking'; other = participants were asked to talk about what they don't like about their body and physical appearance.
GP = general population; SCD = Subjective Cognitive Decline.
H = hypothetical; PR = potentially real.
* Effect size was one of multiple statistically-dependent effect sizes within the study and therefore would have been averaged with those effect sizes for the overall analysis. # Effect size was one of multiple independent subgroups within that study and therefore would have been combined with those effect sizes for the overall analysis.

Re-categorization of control type and participant characteristics. When other moderators were poorly distributed across studies, we were able to re-categorize them to improve that distribution. This was the case for the control-type ('NT' k = 16; 'EPT' k = 6; 'episodic present/recent thinking' k = 8; 'semantic thinking' k = 5; events from a story, k = 5) and participant characteristics ('general population' k = 26; 'alcoholic-dependence' k = 2; nicotine-dependence k = 3; 'obese/overweight' k = 7; 'other' = 4) variables.

We re-categorized control type to create three categories: NT, episodic thinking, semantic/non-personal. The episodic thinking category was created by merging the EPT and episodic present/recent thinking categories. The semantic/non-personal category was created my merging the semantic thinking and 'events from a story' categories. An advantage of the initial categorization is that the different control conditions are in more specific categories which would provide a greater understanding of the important components of the EFT effect. For example, separating semantic thinking and non-personal episodic thinking (events from a story) allows examination of the relative importance of the personal component of the typical episodic thinking conditions. However, as previously stated, a disadvantage of the initial categorization is that some categories are under-represented. Given the advantages and disadvantages of each, results both before and after this re-categorization process are reported.

We dichotomized the participant characteristics categories to create one category of studies drawing from the general population, and another category of studies in which participants might be expected to have steeper delay discounting than participants drawn from the general population (initially coded categories: alcohol-dependence, nicotine-dependence, obese/overweight). EFT might have a larger effect for participants with steeper discounting at baseline because a larger change is possible, and is a particularly relevant intervention for such participants. We also placed studies with participant groups initially coded as "other" into one of these two categories. Participants in the 'other' category were problem gamblers, at-risk drinkers, a lean weight, or vaguely described in the study as "a highly impulsive sample." The problem gamblers and "highly impulsive sample" were placed into the 'impulsive discounting' category, and the 'lean' and 'at-risk drinkers' participants were placed into the general population/ 'less impulsive' category. We put the 'at-risk drinkers' in the general population category because we do not know that this group exhibits steeper discounting. However, results did not change when this group was placed in the 'impulsive discounting' category. Results before re-categorization address potential differences among specific personal characteristics, whereas results after re-categorization address potential differences between participants with steeper discounting in general vs. the general population (while also providing better represented categories across studies). Thus, results before and after re-categorization are reported.

Publication bias. We examined 2 indices of publication bias. Funnel plots of effect size and precision (the inverse of the standard error) were examined via the one-tailed Egger's regression test (Egger, Smith, Schneider, & Minder, 1997), where the standardized effect size (effect size divided by standard error) is regressed on precision. A significant Egger's regression test indicates potential study bias. Secondly, we used Duval and Tweedie's (1998) Trim and Fill procedure which estimates the number of missing studies that may exist in a meta-analysis based on the funnel plot. During this iterative process, the most extreme small studies are removed and the average effect size is recomputed until the plot is symmetric (Borenstein et al., 2009).

Results

Overall effects

EFT had a statistically significant, medium effect on delay discounting across studies, Hedges' $g = 0.48$, 95% CI (0.39, 0.58), $p < 0.001$, $k = 40$. The forest plot is presented in Figure 3.2, and Table 3.2 shows the coded effect sizes and characteristics of the included studies. We found significant heterogeneity across studies, $Q = 73.20$, $p = 0.001$. The I^2 value (46.72) indicates that 46.72% of the observed variability is due to true differences between studies, and may potentially be explained by moderators.

Study name	Hedges' *g*	Variance	Lower limit	Upper limit	p-value
	Statistics for each study				
Banes (2016)	0.168	0.049	-0.264	0.599	0.447
Benoit et al. (2011)	0.318	0.109	-0.328	0.964	0.334
Bromberg et al. (2017)	0.213	0.025	-0.094	0.520	0.174
Bulley and Gullo (2017)	0.554	0.033	0.198	0.910	0.002
Bully et al. (2018)	0.626	0.021	0.342	0.910	0.000
Chiou and Wu (2017)	0.598	0.073	0.070	1.126	0.026
Daniel et al. (2013a)	0.270	0.024	-0.033	0.572	0.081
Daniel et al. (2013b)	1.429	0.219	0.511	2.346	0.002
Daniel et al. (2015)	1.062	0.116	0.394	1.730	0.002
Daniel et al. (2016)	0.378	0.079	-0.172	0.927	0.178
Dassen et al. (2016)	0.483	0.046	0.063	0.903	0.024
Gutsin et al. (2018)	0.450	0.069	-0.063	0.963	0.086
Hu et al. (2017a)	0.208	0.052	-0.237	0.654	0.359
Hu et al. (2017b)	0.097	0.046	-0.323	0.517	0.651
Kakoschke et al. (2018)	0.216	0.106	-0.422	0.854	0.507
Kim et al. (2013.1)	0.556	0.061	0.072	1.040	0.024
Kim et al. (2013.2)	0.639	0.053	0.188	1.090	0.006
Lempert et al. (2012)	1.216	0.089	0.631	1.801	0.000
Lin and Epstein (2014)	0.448	0.048	0.019	0.877	0.041
Liu et al. (2013.1)	0.599	0.040	0.207	0.991	0.003
Mok (2017)	1.483	0.074	0.950	2.016	0.000
O'Donnell et al. (2017)	0.795	0.044	0.382	1.208	0.000
Olsen (2017)	0.614	0.034	0.253	0.975	0.001
Olsen (2018)	0.535	0.014	0.305	0.764	0.000
Palombo et al. (2015)	0.526	0.109	-0.121	1.173	0.111
Peters and Büchel (2010.1)	0.126	0.036	-0.246	0.498	0.507
Peters and Büchel (2010.2)	0.660	0.091	0.069	1.251	0.029
Seidman et al. (2018)	0.906	0.185	0.064	1.748	0.035
Snider et al. (2016)	0.912	0.093	0.314	1.510	0.003
Snider et al. (2018)	0.304	0.027	-0.018	0.626	0.064
Stein et al (2017)	0.407	0.032	0.055	0.760	0.024
Stein et al (2018)	0.419	0.036	0.047	0.791	0.027
Stein et al. (2016)	0.667	0.107	0.026	1.308	0.041
Sze et al. (2017)	0.629	0.031	0.283	0.974	0.000
Wiehler et al. (2017)	0.061	0.017	-0.194	0.317	0.638
Wu et al. (2017.1)	0.746	0.074	0.212	1.279	0.006
Wu et al. (2017.2)	0.679	0.074	0.148	1.210	0.012
Yi et al. (2017.2)	0.273	0.026	-0.043	0.589	0.090
Yi et al. (2017.3)	0.090	0.022	-0.201	0.381	0.544
Zhang et al. (2018.1)	0.492	0.028	0.164	0.820	0.003
OVERALL EFFECT	0.505	0.003	0.406	0.603	0.000

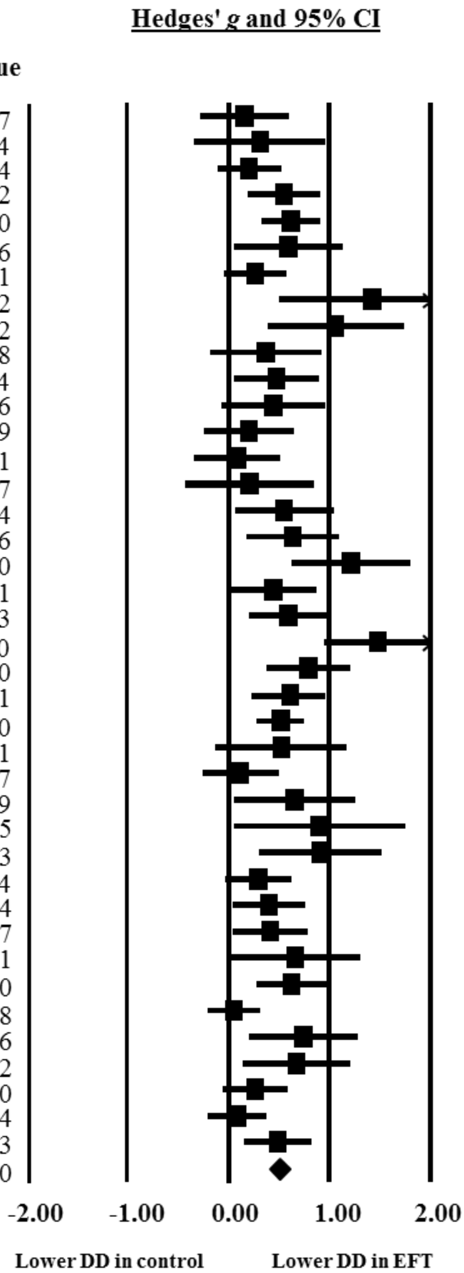

Figure 3.2. Forest plot showing effect sizes (Hedges' *g*) and 95% confidence intervals (CI) for each independent study. Each square represents the effect size of one study, and the horizontal line represents the CI of that effect size. The diamond represents the weighted average effect size across all studies. Effect sizes to the left of zero reflect lower delay discounting (DD) in the control condition, whereas effect sizes to the right of zero reflect lower delay discounting in the EFT condition. Statistics are also shown for each independent study, including Hedges' *g*, variance, 95% CIs and *p*-values. Study letters refer to separate studies written by the same first author within the same year, and study decimal points refer to separate experiments within the same study.

Table 3.2.
Coded effect sizes and characteristics of studies included in the meta-analysis.

Study name	Published	Design	DRM	Control	Participant type	Average age (years)	Reward	Delayed amount (converted to US$)	Longest delay (days)	Hedges' *g*
Banes (2016)	No	BS	AUC	EPT & events from a travel blog	At-risk drinkers	19.92	H	US$100	1825	0.168*
Benoit et al. (2011)	Yes	WS	%LLR & RI	ST	GP	27.30	H	£65 (85.47)	360	0.318*
Bromberg et al. (2017)	Yes	WS	AUC & k	NT	GP	14.50	PR	€49.7 (56.79)	180	0.213*
Bulley & Gullo (2017)	Yes	WS	AUC	Events from a story	GP	20.67	H	AU$10 (7.16)	365	0.554
Bulley et al. (2019)	No	BS	k	ST	GP	19.72	H	AU$85 (60.84)	186	0.626
Chiou & Wu (2017)	Yes	BS	k	NT & ST	smokers	31.40	H	NT$240 (7.79)	365	0.598*
Daniel et al. (2013a)	Yes	WS	AUC	Events from a story	Ob/ow & lean	24.91	H	US$10 & $100	730	0.270*#

Study name	Published	Design	DRM	Control	Participant type	Average age (years)	Reward	Delayed amount (converted to US$)	Longest delay (days)	Hedges' *g*
Daniel et al. (2013b)	Yes	BS	AUC	Events from a story	Ob/ow	26.43	H	US$10 & $100	730	1.429*
Daniel et al. (2015)	Yes	BS	AUC	ERT	Ob/ow	12.23	H	US$50	180	1.062
Daniel et al. (2016)	Yes	BS	AUC	EPT & ERT	GP	26.07	H	US$50	180	0.378*
Dassen et al. (2016)	Yes	BS	*k*	EPT	GP	20.40	H	€85 (97.14)	186	0.483#
Gutsin et al. (2018)	No	BS	*k*	ERT	Ob/ow	60.9	H	US$1,000	N/A	0.450*
Hu et al. (2017a)	Yes	WS	AUC, *k* & % LLR	NT	GP	24	H	€200 (228.56)	365	0.208*
Hu et al. (2017b)	Yes	WS	AUC & *k*	NT	GP	66.49	H	€200 (228.56)	365	0.097*
Kakoschke et al. (2018)	Yes	BS	AUC	NT	Ob/ow	26.93	H	N/A	N/A	0.216

Study name	Published	Design	DRM	Control	Participant type	Average age (years)	Reward	Delayed amount (converted to US$)	Longest delay (days)	Hedges' *g*
Kim et al. (2013.1)	Yes	BS	IPs	NT	GP	25.47	H	£500 (657.66)	365	0.556
Kim et al. (2013.2)	Yes	BS	IPs	NT	GP	23.60	H	£500 (657.66)	365	0.639
Lempert et al. (2012)	Yes	BS	%SSR	NT	GP	20.46	PR	US$10	365	1.216
Lin & Epstein (2014)	Yes	BS	*k*	EPrT	GP	40.87	H	US$85	186	0.448
Liu et al. (2013.1)	Yes	WS	%SSR	NT	GP	20.62	PR	CNY67.50 (10.01)	30	0.599
Mok (2017)	No	WS	AUC	NT	GP	20.24	H	CAD$100 & $2,000 (75.4 & 1507.74)	3,650	1.483*
O'Donnell et al. (2017)	Yes	BS	AUC	EPT	GP	22.25	H	US$100	730	0.795#
Olsen (2017)	No	WS	AUC	NT	GP	N/A	H	NZ$1,000 (683.70)	391	0.614

Study name	Published	Design	DRM	Control	Participant type	Average age (years)	Reward	Delayed amount (converted to US$)	Longest delay (days)	Hedges' *g*
Olsen (2018)	No	WS	AUC	NT, EPT & ST	GP	N/A	H	NZ$1,000 (683.70)	391	0.535*
Palombo et al. (2015)	Yes	WS	RRI	NT	GP	65	H	US$58	730	0.526
Peters & Büchel (2010.1)	Yes	WS	*k*	NT	GP	25.4	PR	€80 (91.42)	176	0.126
Peters & Büchel (2010.2)	Yes	WS	*k*	NT	GP	N/A	PR	€80 (91.42)	176	0.660
Seidman et al. (2018)	No	BS	AUC	ERT	"highly impulsive"	29.31	H	N/A	N/A	0.906*
Snider et al. (2016)	Yes	BS	AUC & IPs	ERT	alcoholics	41.15	H	US$100	365	0.912*
Snider et al. (2018)	Yes	WS	*k*	NT	alcoholics	42.45	H	US$1,000	365	0.304
Stein et al. (2017)	Yes	BS	*k*	EPT	Ob/ow	35.80	H	US$100	9,125	0.407*#

Study name	Published	Design	DRM	Control	Participant type	Average age (years)	Reward	Delayed amount (converted to US$)	Longest delay (days)	Hedges' *g*
Stein et al. (2018)	Yes	BS	*k*	EPT	smokers	36.46	H	US$100	9,125	0.419
Stein et al. (2016)	Yes	BS	AUC	ERT	smokers	39.28	H	US$1,000	365	0.667
Sze et al. (2017)	Yes	BS	AUC	EPT & NT	Ob/ow	37.93	H	US$100	365	0.629*
Wiehler et al. (2017)	Yes	WS	*k*	NT	GP & PG	29.08	PR	€99.50 (113.69)	200	0.061#
Wu et al. (2017.1)	Yes	BS	*k*	NT & ST	GP	20.90	PR	NT$240 (7.79)	365	0.746*
Wu et al. (2017.2)	Yes	BS	*k*	Events from a travel blog & ST	GP	20.20	PR	NT$240 (7.79)	365	0.679*
Yi et al. (2017.2)	Yes	WS	*k*	EprT	GP	20.38	H	US$100	1,825	0.273
Yi et al. (2017.3)	Yes	WS	*k*	EprT	GP	19.55	H	US$100	1,825	0.090

Study name	Published	Design	DRM	Control	Participant type	Average age (years)	Reward	Delayed amount (converted to US$)	Longest delay (days)	Hedges' *g*
Zhang et al. (2018.1)	Yes	WS	%SSR	NT	GP	21.53	PR	CNY172 (25.51)	120	0.492

Note. Olsen (2017) is data from Experiment 1 of the current thesis, and Olsen (2018) is data from Experiment 3A of the current thesis (see Chapter 5).
BS = between-subjects; WS = within-subjects.
DRM = discounting rate measure; AUC = area under the discounting curve; RI = reward index; %LLR = percent (or proportion) larger, later rewards chosen; IPs = indifference points; %SSR = percent (or proportion) smaller, sooner rewards chosen; RRI = re-calibrated reward index (difference between actual accumulated reward and minimum accumulated reward possible divided by the difference between the maximum possible and minimum possible accumulated rewards).
EPT = episodic past thinking; ST = semantic thinking; NT = 'no thinking'; ERT = episodic recent thinking; EPrT = episodic present thinking.
GP = general population; Ob/ow = obese/overweight; PG = problem gamblers.
H = hypothetical; PR = potentially real.
N/A = information unknown.
[*] Effect sizes were averaged across multiple DRMs, controls or delayed amounts. [#] Effect sizes were combined across independent subgroups within a study.

Publication bias

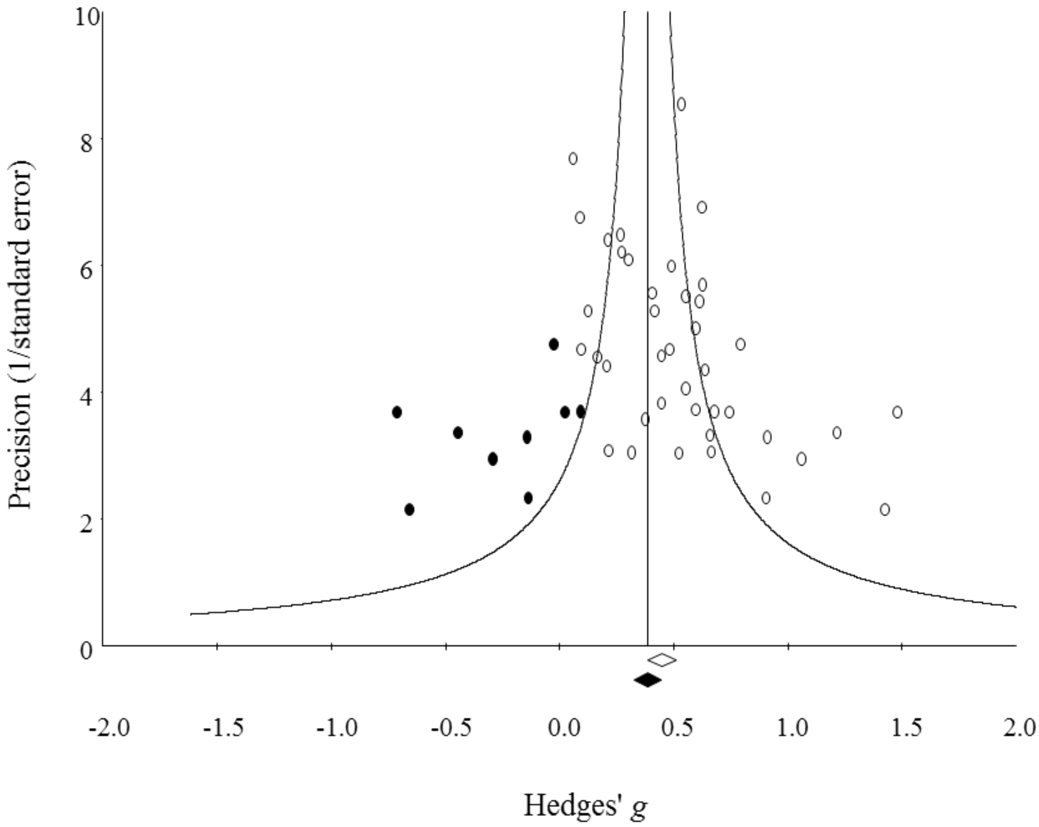

Figure 3.3. Funnel plot of precision by Hedges' *g*. Open circles represent effect sizes from studies included in the meta-analysis, and the open diamond represents the weighted average effect size of those studies. Each black circle represents hypothetically missing studies computed using Duval and Tweedie's (1998) Trim and Fill Test, and the black diamond represents the recomputed average effect size based on these additional studies.

Figure 3.3 shows a funnel plot of each effect size plotted as a function of its precision. The distribution of open circles (effect sizes from studies included in the meta-analysis) in Figure 3.3 provide evidence of publication bias, as studies with small sample sizes and small (or negative) effect sizes are missing. Using Duval and Tweedie's (1998) Trim and Fill approach, 9 hypothetically missing studies were added (black circles in Figure 3.3), resulting in a smaller (although still significant) average effect size, Hedges' $g = 0.38$, 95% *CI* (0.32, 0.44), than in the initial overall analysis with no trimmed studies, Hedges' $g = 0.48$, 95% *CI* (0.39, 0.58). The presence of publication bias was further supported by the significant Egger's regression test, $B0 = 2.21$, $t\ (38) = 3.49$, $p = 0.0006$, indicating that small studies are associated with large effect sizes.

Moderator analyses

Design. Design was a significant moderator of the effect sizes, *Q* (1) = 7.19, *p* = 0.007, such that effect sizes were larger in between-subjects design studies (*g* = 0.60, 95% *CI* [0.48, 0.72], *k* = 22) than in within-subjects design studies (*g* = 0.37, 95% *CI* [0.26, 0.49], *k* = 18).

Discounting rate measure. Discounting rate measure was a significant moderator of the effect sizes, *Q* (2) = 6.27, *p* = 0.044, such that effect sizes were larger when researchers used AUC (*g* = 0.57, 95% *CI* [0.43, 0.71], *k* = 15) than when researchers used *k* (*g* = 0.36, 95% *CI* [0.24, 0.49], *k* = 16). The 'other' category contained studies with an overall effect size that was more comparable to AUC than to *k* (*g* = 0.59, 95% *CI* [0.40, 0.78], *k* = 9).

Control. There was no significant difference in effect size among studies with NT (*g* = 0.45, 95% *CI* [0.30, 0.60], *k* = 16), EPT (*g* = 0.52, 95% *CI* [0.28, 0.75], *k* = 6), episodic present/recent thinking (*g* = 0.48, 95% *CI* [0.26, 0.71], *k* = 8), semantic thinking (*g* = 0.58, 95% *CI* [0.32, 0.83], *k* = 5), or events from a story (*g* = 0.49, 95% *CI* [0.22, 0.76], *k* = 5) control conditions, *Q* (4) = 0.77, *p* = 0.942. Following re-categorization, there was no significant difference in effect size among studies with semantic thinking/non-personal (*g* = 0.56, 95% *CI* [0.35, 0.78], *k* = 8), NT (*g* = 0.45, 95% *CI* [0.30, 0.59], *k* = 16), or episodic thinking (*g* = 0.48, 95% *CI* [0.34, 0.63], *k* = 16) control conditions, *Q* (2) = 0.73, *p* = 0.694.

Reward. There was no significant difference in effect size between studies with hypothetical rewards (*g* = 0.49, 95% *CI* [0.39, 0.60], *k* = 29) and potentially-real rewards (*g* = 0.45, 95% *CI* [0.26, 0.64], *k* = 9), *Q* (1) = 0.16, *p* = 0.687. However, most studies used hypothetical rewards, and thus potentially-real rewards were under-represented. Two studies did not report the reward type (Gutsin et al., 2018; Seidman, Hollis-Hansen, Aref, Hrynyk, & Epstein, 2018).

Participant characteristics. There was no significant difference in effect sizes among studies with participants from the general population (*g* = 0.48, 95% *CI* [0.37, 0.59], *k* = 26), alcohol-dependence (*g* = 0.51, 95% *CI* [0.10, 0.91], *k* = 2), nicotine-dependence (*g* = 0.53, 95% *CI* [0.17, 0.89], *k* = 3), characterized as obese/overweight (*g* = 0.55, 95% *CI* [0.31, 0.79], *k* = 7), or categorized as 'other' (*g* = 0.24, 95% *CI* [-0.07, 0.54], *k* = 4), *Q* (4) = 2.83, *p* = 0.587. Following re-categorization, there was no significant difference in effect size between the 'impulsive discounting' (*g* = 0.50, 95% *CI* [0.34, 0.67], *k* = 14) and 'general population' categories (*g* = 0.46, 95% *CI* [0.35, 0.57], *k* = 28), *Q* (1) = 0.18, *p* = 0.675.

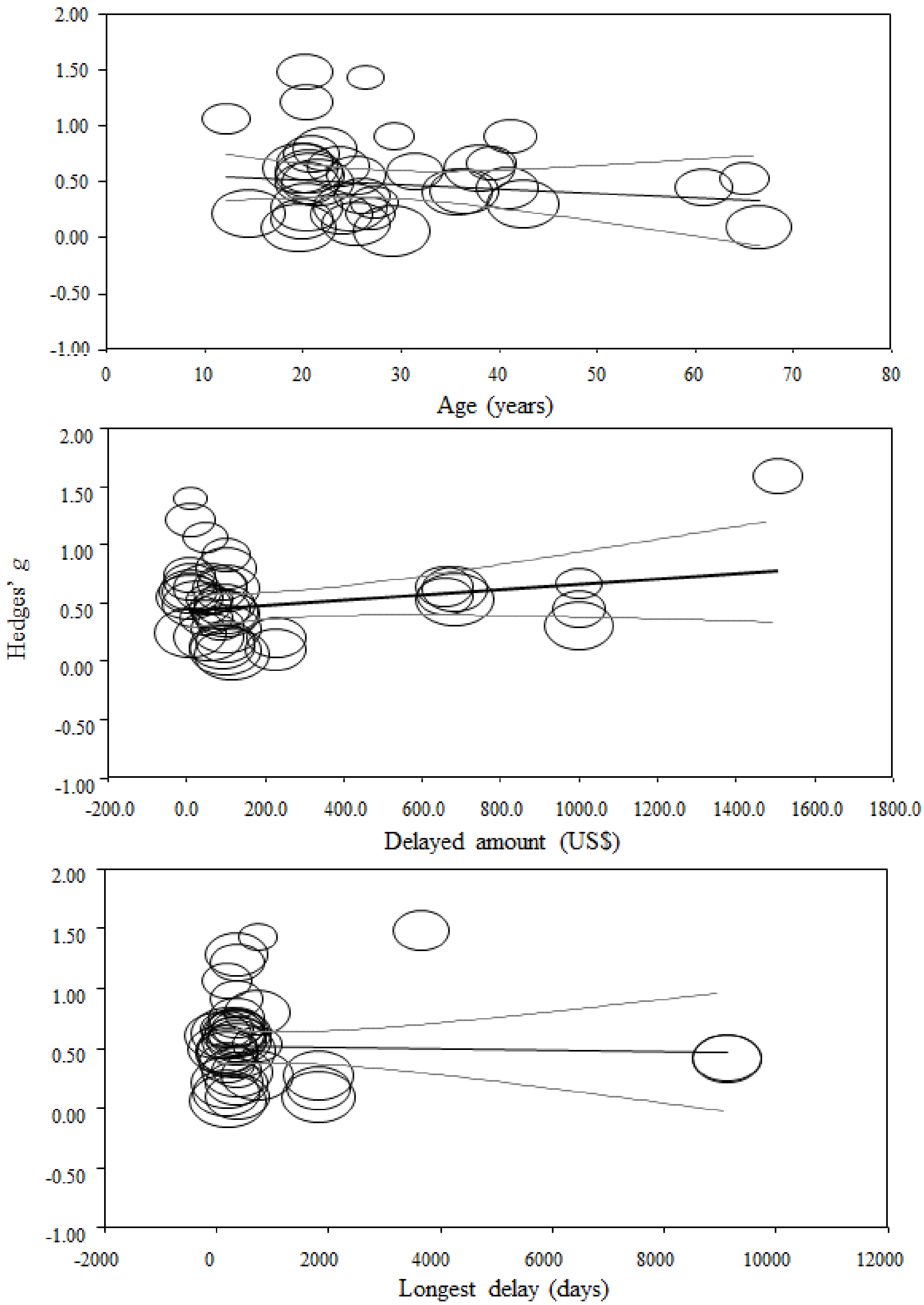

Figure 3.4. Regression of Hedges' *g* on age (top graph), delayed amount (middle graph), and longest delay (bottom graph). For each graph, each circle represents one study, and the bigger the circle, the larger the study weight. The black horizontal line represents the regression line and the grey horizontal lines represent the CI for the regression line.

Age. The mean age of the participants was not a significant predictor of effect sizes, *B* = -0.004, 95% *CI* (-0.01, 0.004), *p* = 0.364, *k* = 37. Figure 3.4 (top graph) shows a scatter plot of the relationship between age and Hedges' *g* to convey the range of mean ages across studies. As can be seen in the graph, most studies recruited participants in their 20s (on average), and participants over 50 years of age were not well-represented across studies. Three studies did not report participants' ages (Olsen, 2017; Olsen, 2018; Peters & Büchel, 2010.2).

Delayed amount. The delayed amount used in the delay discounting task (converted to US$) was not a significant predictor of the effect sizes, *B* = 0.0002, 95% *CI* (-0.00, 0.0005), *p* = 0.092, *k* = 38. One study did not report the delayed amount (Seidman et al., 2018), and one further study was excluded because we could only calculate an effect size combining two delayed amount conditions (Kakoschke, Hawker, Castine, de Courten, & Verdejo Garcia, 2018). Figure 3.4 (middle graph) shows a scatter plot of the relationship between delayed amount and Hedges' *g*. As shown in the graph, ranges above US$200 were not well represented across studies, thus conclusions regarding this moderator are tentative. Results did not change when amounts were re-calculated as a proportion of the country's median household income, *B* = 11.44, 95% *CI* (-2.78, 25.66), *p* = 0.115.

Longest delay. The longest delay used in the delay discounting task was not a significant predictor of the effect sizes, *B* = -0.00, 95% *CI* (-0.00, 0.00), *p* = 0.910, *k* = 37. Figure 3.4 (bottom graph) shows a scatter plot of the relationship between the longest delay and Hedges' *g*. As shown in the graph, the range of delays are poorly distributed across studies; most studies used longest delays of 1 year or less, and studies with delays longer than 2,000 days (5 years) were particularly scarce. Thus, we are unable to make a clear conclusion regarding this moderator. Three studies did not report the longest delay (Seidman et al., 2018; Gutsin et al., 2018; Kakoschke et al., 2018).

Multiple meta-regression

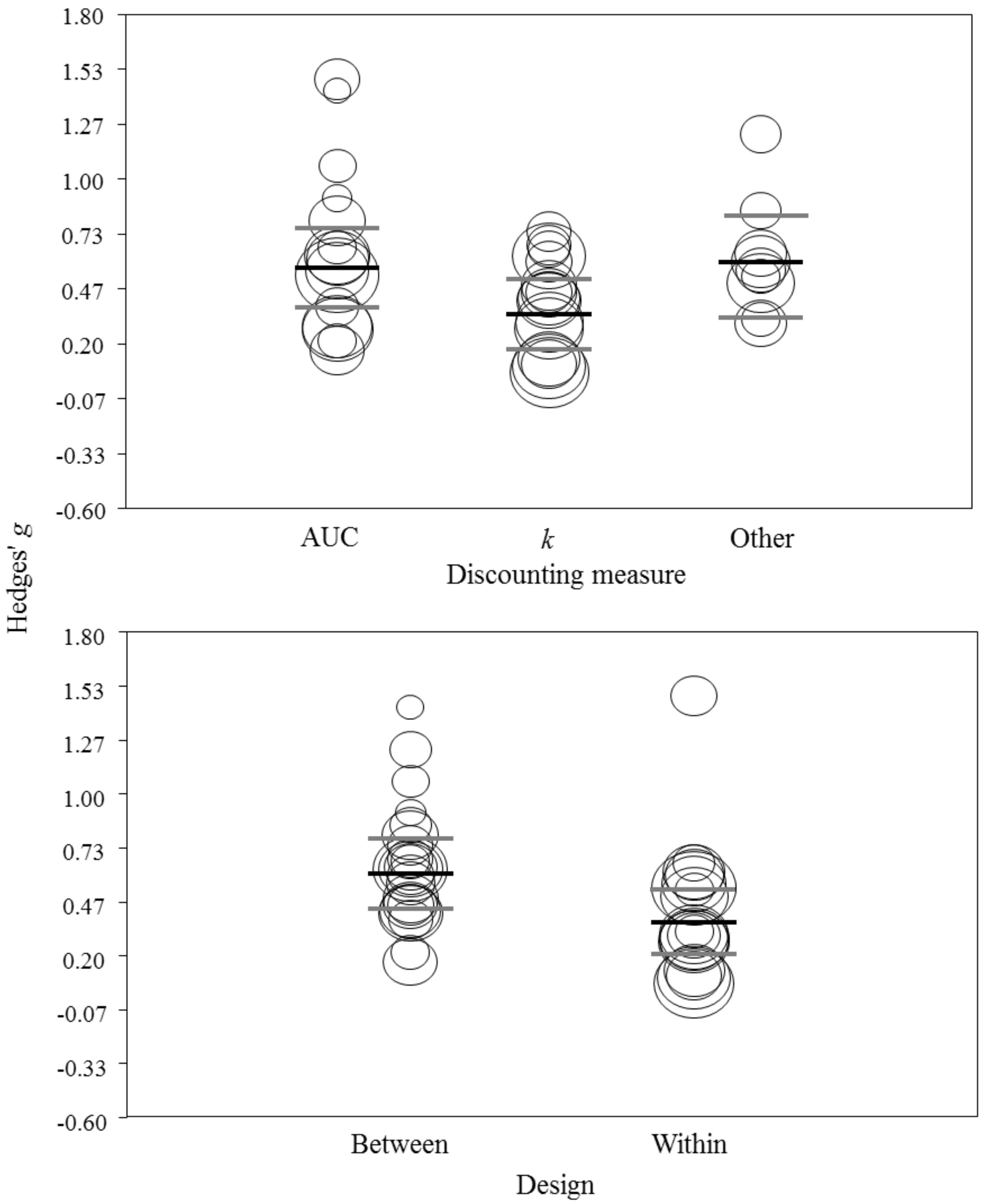

Figure 3.5. Regression of Hedges' *g* on discounting rate measure (top graph) and design (bottom graph) from the multiple meta-regression analysis. Each circle represents one study, and the bigger the circle, the larger the study weight. For each level of each covariate, the black horizontal line represents the regression line and the grey horizontal lines represent the CI for the regression line.

We included both significant moderators (design and discounting rate measure) as covariates in the regression model. For discounting rate measure, we created two dummy variables (‘*k*’ and ‘other’), with ‘AUC’ as the reference group. For design, we created one dummy variable (‘within-subjects’) with ‘between-subjects’ as the reference group. The two moderators were not highly correlated with one another (all *r*s < 0.1).

We found a significant regression model, Q (3) = 18.35, $p = 0.0004$, $R^2 = 0.68$, that explained 68% of the true between-study variance. Both design ($B = -0.24$, 95% *CI* [-0.39, -0.09], $p = 0.001$) and discounting rate measure (Q (2) = 9.80, $p = 0.007$) remained significant moderators. For discounting rate measure type, AUC predicted a significantly larger effect size than *k*, $B = -0.23$, 95% *CI* (-0.39, -0.06), $p = 0.007$, however there was no significant difference between AUC and the ‘other’ category, $B = -0.03$, 95% *CI* (-0.18, 0.24), $p = 0.778$. Figure 3.5 shows scatter plots of the relationship between each covariate and Hedges’ *g*. The test of goodness of fit of the model indicated that there was no significant true between-study variability after accounting for design and discounting rate measure, , Q (36) = 45.27, $I^2 = 20.48$, $p = 0.138$. This analysis is based on all 40 studies.

Discussion

We aimed to quantify the average size of the effect of EFT on delay discounting rate. We also aimed to assess the robustness of the EFT effect across various methodological features and participant characteristics. We found that positive EFT has a significant, medium-sized effect on delay discounting rate. This is consistent with Rung and Madden’s (2018b) meta-analysis, which also found a significant (although small) effect of EFT on delay discounting rate. Publication bias analyses indicated the presence of a file-drawer problem in our meta-analysis, meaning that the average effect size may have been overestimated. However, EFT still had a clear effect on delay discounting rate after we corrected for this bias. We also found significant, moderate levels of true between-study variability in the size of the EFT effect. In contrast, Rung and Madden found little between-study variability. The lack of variability reported by Rung and Madden may reflect their limited sample of studies (their meta-analysis was based on ten studies compared to 40 studies in the current meta-analysis). Moderator analyses revealed that the significant between-study variability found in the current meta-analysis can be substantially explained by the measures used to assess discounting rate and the experimental design used.

Why might within-subjects designs be associated with smaller EFT effects?

The EFT effect was significantly smaller in within-subjects design studies compared to between-subjects design studies. This moderating effect might be explained by the impact of order effects on within subject designs, as previously outlined. Of the within-subjects design studies, 28% used a blocked design where the order of the EFT and control conditions was counterbalanced, meaning that order effects could have increased variability and therefore reduced overall effect sizes. A further 44% of the studies randomized the EFT and control trials, meaning that the EFT trials could have contaminated the control trials, which would also reduce overall effect sizes. Therefore, the majority of the within-subjects design studies (72%) had procedural features that are consistent with the explanation that order effects reduced the size of the EFT effect. Only 28% of the within-subjects design studies are inconsistent with this explanation, where a control block was always presented before an EFT block. However, the absence of a carry-over effect of EFT in Experiment 1 of the current thesis suggests that order effects are an unlikely explanation for the moderating effect of study design (see also Experiments 3A and 3B). Nevertheless, order effects cannot be entirely ruled out, given that at least one study within the meta-analysis reported such order effects (Yi et al., 2017). More researchers should examine whether the EFT effect depends on the order in which the EFT and control conditions are presented.

We previously suggested that demand characteristics would most plausibly produce *larger* effect sizes in studies using within-subjects designs. An effect of demand in the opposite direction cannot be ruled out, however. For example, participants completing two conditions might have attempted to be consistent in their choices in order to be a ‘good participant’, which would reduce the size of the EFT effect in studies using within-subjects designs.

A third explanation might be that within-subjects designs have higher statistical power, and therefore smaller significant effects are more likely to be detected. This explanation implies that between-subjects design studies are more likely to be found in the file-drawer, given that detection of significant, small effects is less likely. To further investigate this possibility, we examined separate funnel plots for each design (see Figure 3.6). Consistent with this explanation, we found stronger evidence of publication bias for the between-subjects design studies (top graph of Figure 3.6), than for the within-subjects design studies (bottom graph of Figure 3.6). That is, there were 5 hypothetically missing between-subjects design studies (identified using Duval & Tweedie’s 1998 Trim and Fill procedure), but no hypothetically missing within-subjects design studies.

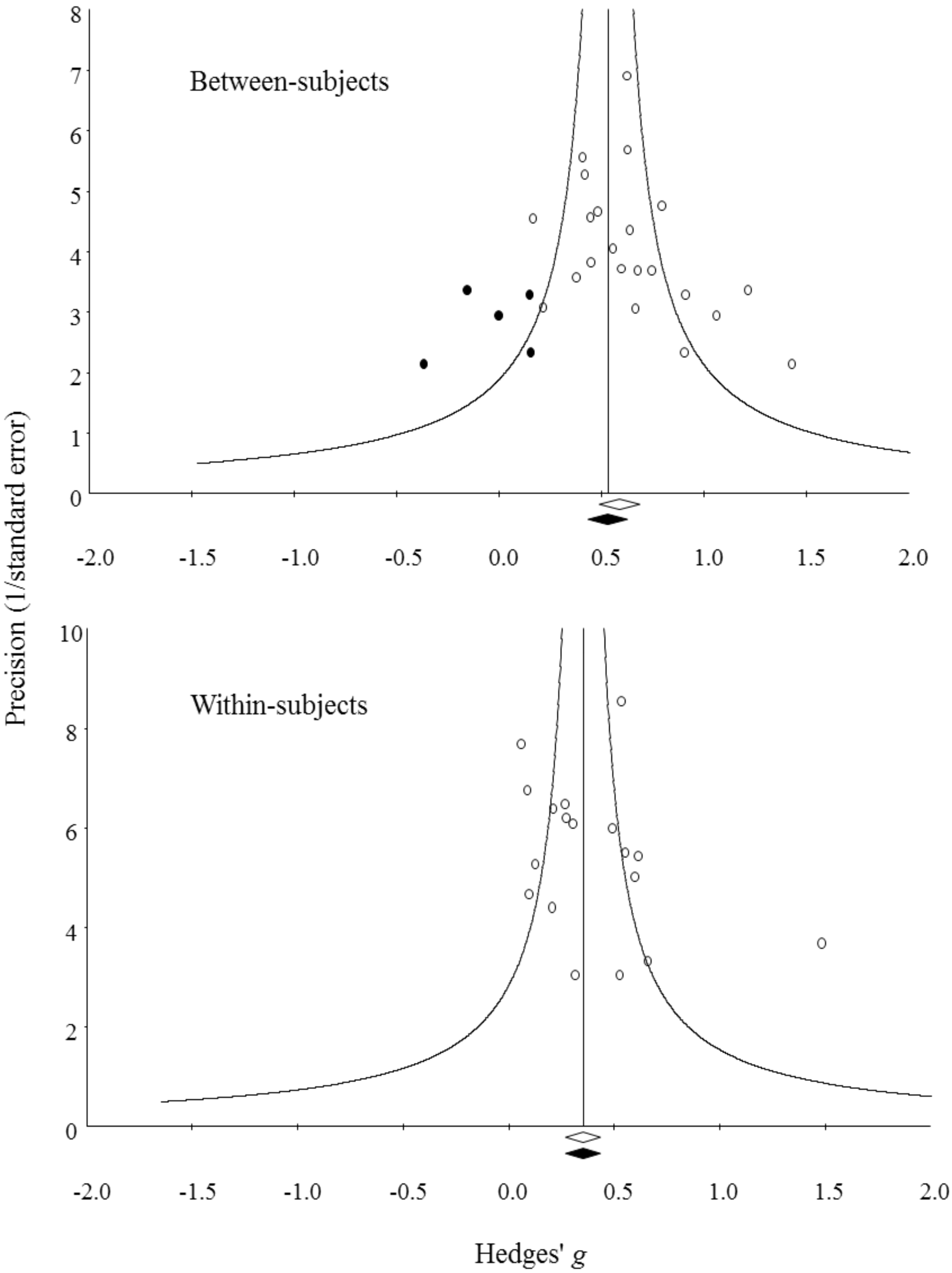

Figure 3.6. Funnel plot of precision by Hedges' *g* for between-subjects design studies (top graph) and within-subjects design studies (bottom graph). Open circles represent effect sizes from studies included in the meta-analysis, and the open diamond represents the weighted average effect size of those studies. Each black circle represents hypothetically missing studies computed using Duval and Tweedie's (1998) Trim and Fill Test, and the black diamond represents the recomputed average effect size based on these additional studies.

Fourthly, it is possible that our within-subjects effect size calculation was unreasonably conservative (Lakens, 2013). We converted all within-subjects Cohen's *d* values into a between-subjects design metric (d_{WC}) using Equation 3.9. However, Lakens (2013) recommends that within-subjects *d* should instead be calculated by dividing the difference in means by the averaged standard deviation of both conditions (ignoring the correlation between conditions), as this is likely to be more comparable to between-subjects Cohen's *d*. As previously stated, we did not calculate within-subjects Cohen's *d* this way because this would not have allowed us to calculate effect sizes for studies where the means and standard deviations were either inappropriate descriptions of the distribution (i.e., non-normal data) or unknown (i.e., only a t-test was reported). Nevertheless, when we re-calculated all within-subjects *d* values using Lakens' recommended equation (and excluded the studies for which this calculation was not possible; $k = 3$), Hedges' *g* was still significantly smaller in within-subjects design studies than in between-subjects design studies, $Q\ (1) = 60.219$, $p < 0.001$. Results from the multiple meta-regression also remained the same (all *p*s < 0.01, except for 'AUC' vs. 'other' $p > 0.05$). Thus, our relatively conservative d_{WC} calculation cannot account for the moderating effect of experimental design.

Why might *k* be associated with smaller EFT effects?

We also found that the size of the EFT effect was smaller in studies where *k* was used as a discounting rate measure, compared to AUC and other, atheoretical measures. AUC might produce larger EFT effects than *k* because longer delays contribute more to the AUC than shorter delays (Borges, Kuang, Milhorn & Yi., 2016), and EFT has been shown to produce larger effects at longer delays (Snider et al., 2016; Stein et al., 2016; Experiment 1 of the current thesis). However, this explanation does not explain why the size of the EFT effect was smaller for *k* than for the other atheoretical measures.

A second explanation could be that participants who are most affected by the EFT manipulation might produce indifference points that are not well described by the hyperbolic model (from which *k* is derived; Equation 1.1). To reiterate, a participant for whom EFT is extremely effective might continuously choose the larger, delayed reward, producing high indifference points across all delays or over many delays. However, the hyperbolic model assumes that indifference points decrease as the delay to the larger amount increases (with the rate of decrease being faster over initial increases in delay) and therefore this participant's data will not be well described by the hyperbolic model, perhaps leading researchers with larger EFT effects to select an alternate discounting rate measure. Thus, for studies where *k* was used as a discounting rate measure, participants may have been less affected by EFT.

Alternatively, participants affected by EFT in the way described above may have been removed, as such a response pattern could be deemed unsystematic (Johnson & Bickel, 2008). Excluding participants with unsystematic discounting based on these criteria is common practice within the delay discounting literature, as such response patterns might reflect participant inattentiveness or misunderstanding of the task. However, this pattern of discounting may also arise as a result of an effective EFT manipulation, rather than inattention, as described above (also see Experiment 1). The removal of such participants from the data set would thus lead to an underestimation of the effect of EFT in the sample. Such researcher decisions might contribute to the association between discounting rate measure and effect size we observed. Consistent with this explanation, nineteen percent of the studies that used *k* as a discounting rate measure reported that they excluded participants who had unsystematic discounting (based on Johnson & Bickel's criteria). A smaller percentage of studies that used AUC (7%) reported that they excluded participants based on these criteria, and no studies that used discounting measures from the 'other' category reported that they used these criteria. Interestingly, when we removed all of these studies ($k = 4$), the significant moderating effect of discounting rate measure become non-significant ($p = 0.098$). However, it is also possible that other procedural decisions that are confounded with discounting rate measure may contribute to this result.

We prefer to use AUC, rather than *k*, when assessing the effects of EFT on delay discounting rate within our lab because it allows inclusion of participants with a wider range of choice patterns. If we had opted to use *k*, we would have had to exclude approximately half of the participants in order to utilize *k* as a more meaningful measure (e.g., see Experiment 3A). However, some unsystematic response patterns might still reflect participant inattentiveness or misunderstanding of the task, as described above. To balance these considerations, we recommend that researchers implement a NT control condition within-subjects in addition to other control conditions of interest. Researchers could then ensure data quality by excluding participants based on Johnson and Bickel's (2008) criteria in the NT condition only, as in Experiment 1. Alternatively, researchers could implement attention check questions to ensure data quality (for an example, see Sze, Stein, Bickel, Paluch, & Epstein, 2017).

Participant characteristics and age

Participant characteristics did not significantly predict the size of the EFT effect. These results are consistent with Daniel et al. (2013a), who found no significant difference in the size of the EFT effect between lean and obese/overweight individuals. However, most studies in the current meta-analysis used participants from the general population, and thus the remaining categories were not well represented across studies. We improved this distribution by dichotomizing studies into ‘general population’ and ‘impulsive discounting’ subgroups. Our reasoning for this re-categorization was that impulsive discounters may show larger EFT effects because a larger decrease in discounting will be possible compared to baseline. The re-categorized participant characteristics variable was not a significant moderator of the effect sizes. This suggests that the EFT effect is indeed robust across participants who exhibit impulsive discounting and the general population, and therefore may be utilized as an effective intervention for a variety of impulsive behaviors.

However, it is still possible that certain subsets of ‘impulsive’ individuals may show smaller EFT effects than others. For example, individuals with alcohol-dependence have reduced working memory capacities compared to healthy controls (Bechara & Martin, 2004), which has been related to reduced EFT effectiveness in one study (Lin & Epstein, 2014). In contrast, working memory is similar between individuals with problem gambling disorder and healthy controls (Lawrence, Luty, Bogdan, Sahakian, & Clark, 2009). Wiehler, Bromberg, and Peters (2015) and Wiehler et al. (2017) also found no evidence of impaired EFT ability in individuals with problem gambling disorder. Such variability within the ‘impulsive discounting’ subgroup of studies could account for the non-significant moderating effect in our meta-analysis. However, there was no significant between-study variability within the ‘impulsive discounting’ subgroup of studies (p = 0.148), suggesting that effect sizes were comparable across the different participant characteristics within the ‘impulsive discounting’ subgroup. Although, it is still possible that higher statistical power was needed to detect such an effect.

We also found no evidence that age was a significant moderator of the size of the EFT effect. This suggests that EFT works equally well between the ages of 12 and 42 years. However, further research is needed to clarify the efficacy of EFT in older adults, as participants over 42 years of age were not well-represented across studies. This under-represented age group is of particular interest, given that Sasse et al. (2017) found no significant effect of EFT on delay discounting rate in healthy participants with an average age of 66.55, and that this was related to age-related cognitive decline. In our meta-analysis, only three studies (Gutsin et al., 2018; Hu et al., 2017b; Palombo et al., 2015; see Table 3.2) used participants similar to the average age of participants used in Sasse et al.'s study. Unfortunately, we were unable to code an effect size from Sasse et al. (2017).

Findings that the EFT effect is robust across various participant characteristics and ages are limited in that moderator analyses within a meta-analysis are observational, rather than experimental, and often have low statistical power (Borenstein et al., 2009). Nevertheless, these results are promising and encourage further research on the development of EFT interventions within clinical settings.

Other non-significant moderators and future directions

The EFT effect was robust across a range of other methodological features. However, these results must be interpreted with caution given that moderators were unevenly distributed across studies and/or covered a restricted range of values (Lipsey, 2003). Nevertheless, the moderators that were unevenly distributed across studies provide a basis for future research. Below we discuss these non-significant moderators in-depth.

Reward. We found that EFT works equally well when hypothetical and potentially-real rewards are used. This suggests that researchers may not need to spend funding on providing outcomes to participants in order to maintain data quality. However, this result should be interpreted with caution, as the majority of studies used hypothetical rewards, and therefore potentially-real rewards were under-represented. Additionally, we were unable to test whether real rewards produce smaller EFT effects because there were no studies in the meta-analysis that used real rewards within their experiments. Future research should assess the effects of EFT on delay discounting rate using real rewards.

Control condition. Understanding the important components of EFT will result in EFT being translated into clinical interventions more efficiently. If changes in discounting are largest between EPT and EFT conditions, then this would suggest that future thinking is the more important component of EFT, given that this is the one component missing from EPT controls. Alternatively, if changes in discounting are largest between semantic future thinking and EFT conditions, then this would suggest that episodic thinking is the more important component of EFT, given that this is the one component missing from semantic future thinking controls. We found no evidence that the type of control condition predicts the size of the EFT effect. This could suggest that episodic thinking and future thinking are equally important components of the EFT effect. However, this conclusion is tentative given that most studies used a NT control condition and active controls were therefore under-represented. We improved this distribution by merging the EPT and episodic present/recent thinking conditions, and by merging the semantic thinking and non-personal (i.e., 'events from a story') conditions. Consistent with the original results, we found no significant difference in effect size among the re-categorized control conditions.

A remaining caveat regarding the initial categorisation is that not all semantic thinking conditions had a future component. Therefore, differences between EFT and semantic non-future thinking may be explained by episodic and/or future thinking, given that both of these components are missing from semantic non-future thinking controls. Combining both semantic future and semantic non-future conditions and comparing this category to EFT therefore makes determining the relative importance of future and episodic thinking difficult. When we re-ran the moderator analysis with semantic non-future thinking conditions either as a separate category or excluded, results were still non-significant (*p*s > 0.05). However, conclusions remain tentative given the limited number of studies that used semantic non-future thinking (k = 2) and semantic future thinking (k = 3) control conditions. Additionally, following re-categorization, the semantic/non-personal category contained a mix of: 1) semantic future thinking, 2) semantic non-future thinking, and 3) episodic, non-future, non-personal thinking (i.e., events from a story). Combining all of these conditions within one category again makes determining the relative importance of future and episodic thinking difficult. The relative importance of the episodic and future thinking components of EFT therefore remains to be determined.

Further discussion of the emotional valence of EFT

The current meta-analysis included only studies that instructed participants to think about positively-valenced future events. Nevertheless, the impact of valence on how effectively EFT promotes self-control warrants further discussion in general. As shown in Table 3.1, most studies that used only negative EFT found medium, negative effect sizes (except for Bulley et al., 2019, who found a medium, positive effect size). Additionally, studies that used only neutral EFT found very small (null) effects. Although Lin and Epstein (2014) also assessed the effects of neutral-only EFT, we were unable to code this effect size, and therefore this study is not in Table 3.1[7]. Interestingly, Lin and Epstein found that neutral EFT decreased delay discounting rate, in contrast to Liu et al. (2013) and Zhang et al. (2018). As also shown in Table 3.1, Parthasarathi et al. (2017) used an EFT manipulation that consisted of both positive and negative components. Participants were asked to focus on a goal that they would like to achieve in the future, and imagined achieving that goal (positive valence) as well as scenarios where they overcame obstacles to those goals (possible negative valence). Interestingly, this resulted in a very small (and negative) effect size.

Discrepancies in the effects of negative and neutral EFT may be due to differences in methodological procedures across studies. A notable difference across all three studies with discrepancies in results is the different type of control conditions used. That is, Liu et al. (2013) and Zhang et al. (2018) compared neutral and negative EFT to a NT control, whereas Lin and Epstein (2014) compared neutral EFT to neutral episodic present thinking. Similar to Lin and Epstein, Bulley et al. (2019) compared negative EFT to thinking about neutral activities, however these did not contain a temporal component. Thus, overall the effect of considering future events of neutral, negative or mixed valence remains to be determined.

Conclusion

EFT had a significant, medium-sized effect on delay discounting rate across studies. The size of the EFT effect was robust across different participant characteristics and ages, again highlighting the therapeutic potential of EFT. When designing EFT studies, researchers should be mindful of the type of discounting rate measure and study design used, as this may affect the size of the EFT effect. The size of the EFT effect was robust across other methodological features, including: 1) type of control condition, 2) longest delay to the larger, delayed reward, 3) amount of the larger, delayed reward, and 4) reward type. However, these latter, non-significant results should be interpreted with caution, given that these variables were not well represented across studies. Further research is warranted in order to form a better understanding of how these methodological features influence EFT efficacy. Experiment 3A in Chapter 5 further addressed the relative importance of the episodic and future thinking components of EFT.

---

[7] We were only able to calculate one effect size from Lin and Epstein (2014) that combined both positive and neutral conditions, based on the available statistics in the paper (see Table 3.2). However, we did not exclude this effect size from the overall meta-analysis given that the authors reported no significant difference in the size of the EFT effect between the positive and neutral conditions.

Chapter 4: Assessing the validity of the report indifference points procedure (Experiment 2)

Chapter 4 details three experiments where we investigated the validity of an alternative procedure to the titrating-amount procedure that could be used to measure delay discounting: The *report indifference points procedure*. In the report indifference points procedure, participants are asked to simply report their indifference points, rather than the researcher calculating these based on several trials. To obtain an indifference point for six different delays, participants would only be asked six questions, in contrast to the 30 questions required with our titrating-amount procedure. This faster procedure would be useful for subsequent experiments where we intended to manipulate multiple conditions within subjects.

In one study that used a report indifference points procedure, participants were asked to use a slider tool to tell the researcher the amount of money received immediately that would make them feel just as good as receiving the money after the specified delay (Johnson et al., 2015). Johnson et al. (2015) found that this particular method produced low levels of unsystematic discounting (less than 9% according to Johnson & Bickel's 2008 criteria). They also found that the procedure produced well-established findings in the discounting literature, including steeper discounting of smaller, delayed rewards than larger, delayed rewards (the magnitude effect; e.g., Baker, Johnson, & Bickel, 2003, Benzion, et al., 1989; Chapman, 1996; Green et al., 1997; Raineri & Rachlin, 1993) and steeper discounting of cigarettes than money in smokers (Baker et al., 2003; Johnson, Bickel, & Baker, 2007; Odum & Baumann, 2007). Johnson et al.'s (2015) findings support the validity of their report indifference points procedure.

Other studies that have used a report indifference points procedure have asked participants to specify the value of the smaller, immediate reward in a blank space, rather than use a slider tool to indicate this value. This is known as the *fill-in-the-blank (FITB)* method (e.g., Chapman, 1996; Read, Frederick, Orsel, & Rahman, 2005; Smith & Hantula, 2008; Weatherly & Derenne, 2011). Smith and Hantula (2008) compared discounting rates between the FITB procedure and a binary-choice procedure. The binary-choice procedure was comparable to the titrating-amount procedure in that participants made a series of choices between two alternatives: a smaller, sooner and a larger, delayed reward. The FITB procedure produced shallower discounting compared to the binary-choice procedure, and participants reported that the FITB procedure was cognitively demanding. This high cognitive load could have resulted in high levels of unsystematic discounting in the FITB procedure (Franco-Watkins, Rickard, & Pashler, 2010). Smith and Hantula did not assess levels of unsystematic discounting for either procedure, although they did report that the hyperbolic model described their data well (mean R^2 = 0.99), indicating that the number of participants demonstrating unsystematic discounting was unlikely to be high. However, Weatherley and Derenne (2011) reported that the hyperbolic model did not describe discounting from the FITB procedure well in their study (mean R^2 = 0.31-0.39), and suggested that this could be due to the cognitively demanding nature of the task.

We used Johnson et al.'s (2015) slider tool method for our report indifference points procedure, given that this has been previously shown to produce low levels of unsystematic discounting. It is likely that the FITB method is more cognitively demanding than the slider tool method because participants are required to generate values themselves, rather than make a choice based on values provided by the researcher. However, Johnson et al.'s slider tool method has not been frequently used in the delay discounting literature, nor has it been directly compared to a well-established binary-choice procedure, such as the titrating-amount procedure. We therefore conducted a series of experiments to determine whether Johnson et al.'s results are replicable and to further validate their report indifference points procedure.

Experiment 2A

In Experiment 2A, we compared levels of unsystematic discounting and discounting rates between the titrating-amount procedure and Johnson et al.'s (2015) report indifference points procedure within subjects. Discounting rates obtained from binary-choice procedures have been shown to predict impulsive behaviours (e.g., Dallery & Raif, 2007; Krishnan-Sarin et al., 2007; Washio et al., 2011; Yoon et al., 2007), indicating that they produce a valid measure of discounting. Thus, evidence to support the validity of the report indifference points procedure would include: 1) similar discounting rates between the titrating-amount and report indifference points procedures (or at least a strong correlation between discounting rates obtained from the two procedures), and 2) similar levels of unsystematic discounting between the two procedures (and within the typical range; 18% or less, Smith et al., 2018). These results would indicate that the report indifference points procedure will be appropriate to use in subsequent experiments of the current thesis. We also considered whether the order in which participants experienced the two procedures affected the results. Participants might have a better understanding of the report indifference points procedure if they experience the titrating-amount procedure first, which would produce discounting rates and levels of unsystematic discounting that are more similar to the titrating-amount procedure. Given that EFT may produce unsystematic discounting as a consequence of the manipulation, an EFT manipulation was not included in the current experiment.

Method

Participants

Participants were 502 first-year psychology students at Victoria University of Wellington recruited in partial fulfilment of a course requirement. No other demographic information was collected. Informed consent was obtained prior to participation.

Materials

Participants completed the following tasks online using Qualtrics Survey Software (Qualtrics, Provo, UT). See Johnson et al. (2015) and Olsen (2016) for evidence that using online data collection for delay discounting research does not reduce data quality.

Report indifference points procedure. Participants read the following instructions (adapted from Johnson et al., 2015):

> For each of the next six questions, imagine the following hypothetical scenario: you are presented with a choice between money now or later, **your job is to use the slider tool to tell us the amount of money that you would like to receive immediately** that would make you feel JUST AS GOOD as you would if you were to receive money after the specified time. Although the scenarios are pretend, we ask that you consider each scenario as if it was real and as if it was the only scenario you would face today.

Participants were asked to answer this question (by selecting a given amount of money using the slider tool) for six different delays (7 days, 14 days, 30 days, 120 days, 210 days, and 365 days). The slider tool ranged from $0 to $1,000, and the delays were presented in an ascending order. Each trial/choice was presented on its own screen. The slider tool was at $1,000 at the start of each trial. Participants clicked a ‘next’ button to move onto the next choice with no option to go back and view their previous choices. Figure 4.1 shows a screenshot of the task, using the 365 days delay as an example.

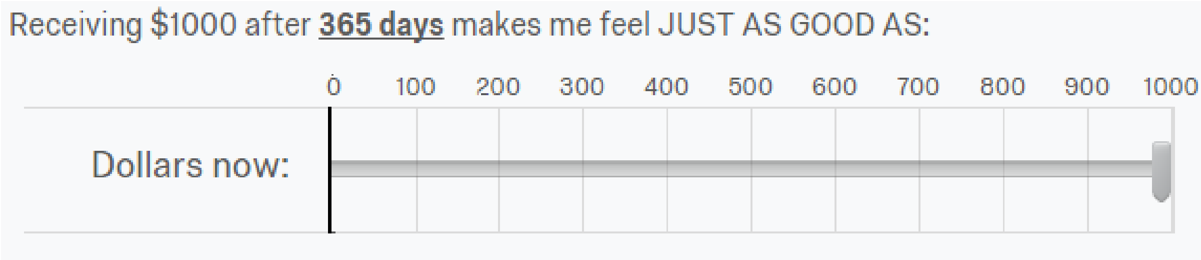

Figure 4.1. Screenshot of one of the six questions participants were asked during the report indifference points task.

Titrating-amount procedure. Participants read the following instructions:

> You will now be asked a set of hypothetical questions that require you to make choices between money now or later. Although the scenarios are pretend, we ask that you consider each scenario as if it were real and as if it was the only scenario you would face today. Please read each question carefully as the amount of money you can receive and when you can receive the money will vary.

The participants were asked to indicate their preference between two hypothetical choices: one reward that was smaller in magnitude and available immediately ($500 now) and a larger reward that was available after a delay (receive $1,000 after X delay). As in the report indifference points task, there were six delays (7 days, 14 days, 30 days, 120 days, 210 days, and 365 days), which were presented in an ascending order. An indifference point was identified for each delay using the titrating-amount procedure, as described in previous chapters of the current thesis. There were five trials for each delay, as in Experiment 1.

Procedure

All participants experienced both tasks in a counterbalanced order. They were then presented with a written debrief on their computer screen which explained the purpose of the experiment. The entire survey took approximately 10 minutes to complete.

Data analyses

In the report indifference points procedure, participants could use the slider tool to report a maximum indifference point of $1,000. However, indifference points could not be higher than $950 in the titrating-amount procedure, given that there were only five titrations. Thus, in order to make AUC more comparable between the two procedures, indifference points in the report indifference points procedure were constrained to a maximum of $950. This constraint required 29% of the indifference points obtained in the report indifference points task to be reduced (from a value between 950 and 1000 to 950). The constraint was applied to at least one indifference point in 56% of participants. Most (63%) of the constraints were applied at the shorter delays (7 days, 14 days, 30 days). The hyperbolic model (Equation 1.1) was then fit to each participants' indifference points for each discounting procedure. We also calculated AUC (Equation 1.2) for each procedure for each participant.

Tests of normality (Shapiro-Wilk and visual inspection of histograms depicting data distributions) showed that k and AUC values were not normally distributed (all ps < 0.05). Although Shapiro-Wilk tests are unreliable with large samples (N >300) because they are too sensitive (Kim, 2013), visual inspection of histograms confirmed high negative kurtosis for AUCs for the report indifference points procedure. There was also a high positive skew and kurtosis for k values for both procedures. Log transformations failed to normalise the k and AUC values, and there is no non-parametric equivalent of a two-way mixed ANOVA. We therefore compared discounting rates between the two procedures using a robust two-way mixed ANOVA (procedure x order), which does not require a normal distribution (Wilcox, 2012). This robust ANOVA uses trimmed means and a bootstrapping procedure. A trimmed mean is when a percentage of the data are removed from each end of the distribution before the mean is calculated. We trimmed 20% of the data as this is commonly used and provides a good estimate of the mean (Wilcox, 2012). A bootstrapping procedure estimates the shape of the sampling distribution by taking repeated samples of the data. We set the number of bootstrapping samples to 2000, as Field (2009) suggested that this will be more than enough to estimate the distribution. We ran the robust ANOVA using R statistical software, using the R function bwtrimbt from the R package WRS (Wilcox, 2012). The bootstrapping procedure does not provide F statistics, only p values. We conducted post hoc tests using the R function ydbt, which is a robust paired samples t test that uses bootstrapping and trimmed means. We used a Bonferroni-corrected alpha of 0.025 when interpreting the post hoc tests, and the reported means are the trimmed means.

We identified the percentage of participants with unsystematic discounting in each procedure based on Johnson and Bickel's (2008) criteria. A problem with the report indifference points procedure was that nearly half of the participants had unsystematic discounting (see Results and Discussion below). We therefore compared discounting rates between the two procedures when unsystematic participants were included as well as when they were excluded[8]. Given that researchers might reasonably choose to analyse these data using either of these approaches, it was important to determine whether this decision affected the results. Given the large body of evidence that indicates that, in general, the value of a reward decreases as a monotonic function of delay (e.g., Green & Myerson, 1996; Green et al., 1999; Madden et al., 2003; Madden et al., 1999; Myerson & Green, 1995; Rachlin et al., 1991), it was assumed that participants with systematic data understood and paid attention to the task, relative to participants with unsystematic discounting. Thus, if the AUC/*k* results are consistent when unsystematic participants are both included and excluded, then this would indicate that a failure to understand the task instructions alone is an unlikely explanation for any differences in discounting between the two procedures.

Results and Discussion

Levels of unsystematic discounting

Only 13% of participants showed unsystematic discounting in the titrating-amount procedure, which is within the typical range (Smith et al., 2018). However, approximately half (49%) of the participants showed unsystematic discounting in the report indifference points procedure, and this was the case regardless of the order in which participants experienced the tasks (titrating-amount first = 50%; report indifference points first = 48%). Table 4.1 shows the percentage of participants with unsystematic discounting based on each of Johnson and Bickel's (2008) criteria. As shown in the table, most participants had unsystematic discounting according to criterion 2 (i.e., the indifference point at the last delay was not less than the indifference point at the first delay by at least 100). Less than half of the participants who failed criterion 2 continuously chose the larger, delayed reward (i.e., had a flat discounting function). Rather, most of these participants in the report indifference points procedure had a higher indifference point at the last delay than at the first delay; the difference between the last delay and the previous delay was not large enough to violate criterion 1 in addition to criterion 2. This pattern of responding in the report indifference points procedure is inconsistent with delay discounting of reinforcer value observed in everyday settings. The percentage of participants exhibiting this pattern of responding is also

[8] If a participant had unsystematic discounting in one or both procedures, then that participant's entire dataset was excluded. That is, they did not contribute to the mean for either procedure.

well above the typical range. These findings therefore indicate that too many people either did not understand the questions or were unable to accurately answer the questions in the report indifference points procedure.

Table 4.1
Percentage of participants with unsystematic discounting (based on Johnson & Bickel's 2008 criteria) in the titrating-amount and report indifference points tasks.

Task	Criterion 1 only	Criterion 2 only	Both criteria
Titrating-amount	1%	11% (flat function = 8%)	0%
Report indifference points	5%	33% (flat function = 12%)	11%

The high percentage of unsystematic discounting found in the report indifferent points procedure is inconsistent with Johnson et al. (2015), who found that less than 9% of participants had unsystematic discounting. However, Rung, Argyle, Siri, and Madden (2018) have recently published research (after the current Experiment 2A data were collected) showing that 47% of participants had unsystematic discounting in the report indifference points procedure, similar to the current study. Rung et al. (2018) also found lower levels of unsystematic discounting in the titrating-amount procedure compared to the report indifference points procedure, however, levels were still higher than the typical range (28%). Rung et al. suggested that the higher levels of unsystematic discounting found in the titrating-amount procedure in their study might be due to the fact that they used university students. This reasoning was based on the meta-analysis by Smith et al. (2018) which showed that unsystematic discounting was higher when participants were recruited from university settings (20.9% vs 15.3% from non-university settings). However, this explanation is inconsistent with the current experiment, as we found low levels of unsystematic discounting that were in the typical range in the titrating-amount procedure using university students.

An alternative explanation that could account for the higher levels of unsystematic discounting in the titrating-amount procedure found in Rung et al. (2018) is that Rung et al. applied a stricter criterion to determine unsystematic discounting than what is described by Johnson and Bickel (2008). Johnson and Bickel's criterion 1 states that discounting is unsystematic if any indifference point (starting with the second delay) is greater than the preceding indifference point by a magnitude *greater than 20%* of the larger, later reward.

This was the criterion used in the current study. However, Rung et al. defined discounting as unsystematic if any indifference point is greater than the previous *by 20% or more* of the larger, later reward. The percentage of unsystematic discounting in the titrating-amount procedure of the current experiment increased from 13% to 19% when we applied Rung et al.'s stricter criterion. Although, Rung et al.'s percentage is still notably higher at 28%.

Discounting rates

The hyperbolic model described participants' data well in the titrating-amount procedure (*Mdn* R^2 = 0.80). However, the hyperbolic model did not describe participants' data well in the report indifference points procedure (*Mdn* R^2 = -0.49), consistent with Weatherly and Derenne (2011). Johnson et al. (2015) reported that the hyperbolic model described discounting in the report indifference points procedure better than an exponential model[9], however, they did not quantify absolute model fits. Even after the unsystematic participants were removed from the current study, over one third of the participants (33%) still had R^2 values for the hyperbolic fits that were less than 0.5 in the report indifference points procedure. Examination of the residuals for the systematic participants also showed that the model overestimated value at the short delays. This was likely due to the fact that the indifference points were constrained to $950 whereas the model assumes indifference points can be as high as $1,000 and the model predicts a subjective value of $1,000 (A) where delay = 0. Consistent with this explanation, it was primarily at shorter delays participants had indifference points of $950 following implementation of the necessary constraint (as previously described). However, the model also underestimated value at the long delays (see Appendix E). We therefore used AUC as the primary discounting measure, however results did not change when *k* was used as a discounting measure (see Appendix F).

[9] In contrast to an exponential model, the hyperbolic model describes a larger proportional decrease in subjective value at shorter delays and a smaller proportional decrease at longer delays (Myerson & Green, 1995). Within the general discounting literature, the hyperbolic model tends to provide a better fit to the indifference points than the exponential model (e.g., Green & Myerson, 1996; Green et al., 1999; Madden et al., 2003; Madden et al., 1999; Myerson & Green, 1995; Rachlin et al., 1991).

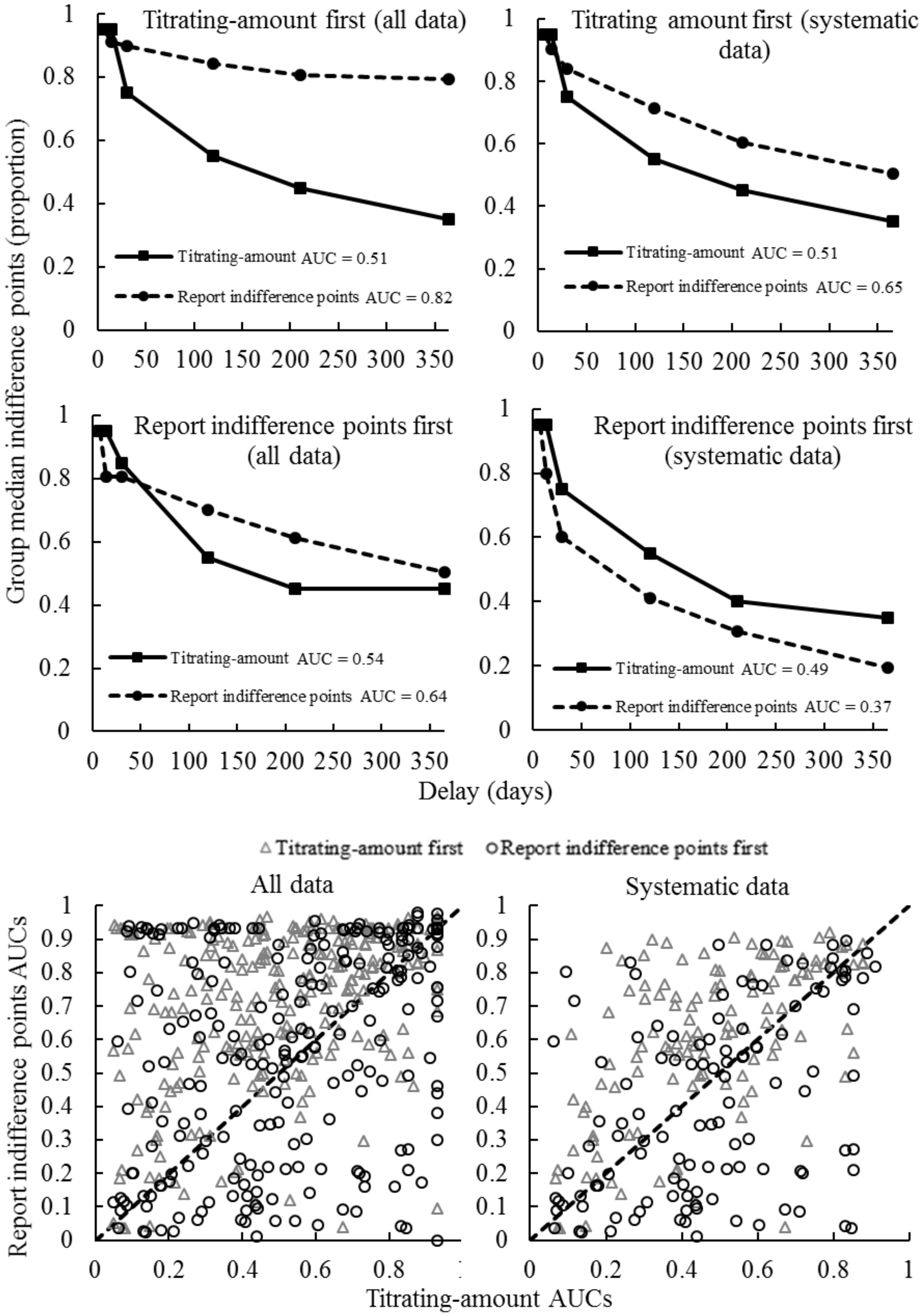

Figure 4.2. The effect of task order based on data from all participants (left panel) and based on data from systematic participants only (right panel). The top line graphs depict the group AUCs for the titrating-amount (squares and solid line) and report indifference points (circles and dashed line) procedures calculated from participants who experienced the titrating-amount procedure first. The middle line graphs depict the same data based on participants

who experienced the report indifference points procedure first. The bottom graphs depict the individual data where AUC values for the titrating-amount procedure are plotted against AUC values for the report indifference points procedure. Each plotted circle/triangle represents one individual; the black circles represent individuals who experienced the report indifference points procedure first, whereas the grey triangles represent individuals who experienced the titrating-amount procedure first. Plotted circles/triangles above the dashed line represent participants with higher AUCs in the report indifference points procedure, whereas plotted circles/triangles below the dashed line represent participants with higher AUCs in the titrating-amount procedure.

The line graphs in Figure 4.2 show the group median indifference points as a function of delay for participants who experienced the titrating-amount procedure first (top panel graphs) and for participants who experienced the report indifference points procedure first (middle panel graphs). The data points are joined by lines to depict the AUCs. The bottom panel graphs represent the individual data. The left panel graphs show the AUCs based on all participants' data. The robust two-way mixed ANOVA showed significant main effects of procedure ($p < 0.001$) and order ($p = 0.042$) and a significant interaction between procedure and order ($p < 0.001$). Post hoc tests revealed that the AUCs were significantly higher in the report indifference points procedure when the titrating-amount procedure was experienced first (report indifference points $M = 0.77$; titration $M = 0.51$; $p < 0.001$). However, there was no significant difference in the AUCs when the report indifference points procedure was experienced first (report indifference points $M = 0.62$; titration $M = 0.56$; $p = 0.065$). This order effect was driven by the report indifference points procedure, as only the AUCs from the report indifference points procedure were significantly different between the two order conditions ($p < 0.001$). There was no significant difference in the AUCs between the two orders conditions for the titrating-amount procedure ($p = 0.104$)[10].

When we included only data from systematic participants in the analysis (see right panel graphs in Figure 4.2), the robust two-way mixed ANOVA showed significant main effects of procedure ($p = 0.044$) and order ($p < 0.001$) and a significant interaction between procedure and order ($p < 0.001$). The pattern of results evident in the follow-up tests differed from when all participants were included in the analysis. Although AUCs were significantly higher in the report indifference points procedure when the titrating-amount procedure was experienced first (report indifference points $M = 0.65$; titration $M = 0.48$; $p < 0.001$), AUCs

[10] The post hoc between-group comparisons were tested using the R function yuenbt, which is a robust independent samples t test that uses bootstrapping and trimmed means.

were significantly *lower* in the report indifference points procedure when the report indifference points procedure was experienced first (report indifference points $M = 0.39$; titration $M = 0.48$; $p = 0.010$). These results show that the systematic participants were more impulsive on the task that they experienced first. However, this order effect was again primarily driven by the report indifference points procedure, as only the AUCs from the report indifference points procedure were significantly different between the two order conditions ($p < 0.001$). There was no significant difference in the AUCs between the two order conditions for the titrating-amount procedure ($p = 0.989$).

If more participants had flat discounting functions in the report indifference points procedure in the 'report indifference points procedure first' order condition compared to in the 'titrating-amount procedure first' order condition, then this could explain the different results when unsystematic participants were included vs. excluded. This is because those individuals' discounting functions would be identified as unsystematic, and excluding those participants would reduce the mean AUC in the report indifference points procedure. However, the percentage of participants who demonstrated flat discounting functions in the report indifference points procedure was similar between the two groups (report indifference points first = 13%; titrating-amount first = 11%). Thus, this cannot explain the inconsistent results.

Smith and Hantula (2008) found that discounting was shallower in the FITB procedure compared to a binary-choice procedure using a within-subjects design. However, Rung et al. (2018) found no difference in discounting rate between the titrating-amount and report indifference points procedures using a between-subjects design. Consistent with Smith and Hantula, we found that discounting was shallower in the report indifference points procedure than in the titrating-amount procedure overall, regardless of whether the unsystematic participants were included or excluded (*p*s < 0.05). However, Smith and Hantula did not consider the effect of task order in their experiment. Taken together, these findings indicate that the difference in discounting between the two procedures is specific to a within-subjects design, and that the differences are likely driven by the fact that the report indifference points procedure is less robust in that it is affected by condition order whereas the titrating-amount procedure is not.

Correlations

Discounting rates obtained from the report indifference points procedure and the titrating-amount procedure were significantly correlated according to both AUC and *k*, regardless of whether unsystematic participants were included or excluded (all *p*s < 0.001). However, the size of the correlation coefficients ranged from 0.33 to 0.47, which are notably weaker than the correlations that researchers have previously reported between other delay discounting procedures (0.67-0.91; Epstein et al., 2003; Holt, Green & Myerson., 2012; Koffarnus & Bickel, 2014). The size of the correlation is therefore not as strong as would be expected in two measures of the same underlying construct. The order effect we observed might have contributed to the weaker correlation we observed by adding an additional source of variability to discounting as measured by the report indifference points procedure.

Conclusion

Findings that nearly half of the participants' data from the report indifference points procedure were unsystematic shows that this procedure will not be appropriate to use for future studies of the current thesis, as it indicates that too many people did not understand the questions or were unable to accurately answer the questions. We also found that participants' discounting rates in the report indifference points procedure were affected by the order in which they completed the tasks. Although this was the case regardless of whether unsystematic participants were included or excluded, the precise details of the order effects obtained did change. Researchers should be aware of the fact that choosing whether to include or exclude unsystematic participants can affect their results. Finally, we found that the report indifference points procedure and the well-established titrating-amount procedure were not strongly correlated, indicating that they likely measure different constructs.

In Johnson et al.'s (2015) study, participants read the instructions for the report indifference points procedure and then completed a quiz to assess their understanding of those instructions. If the question in the quiz was answered incorrectly, then the participant was asked to try again until they answered correctly before beginning the report indifference points task. We did not incorporate this quiz into our report indifference points procedure in Experiment 2A, nor did Rung et al. (2018). It is possible that this quiz helped participants to better understand the questions that they were asked, which could account for the lower levels of unsystematic discounting found by Johnson et al.

Experiment 2B

In Experiment 2B, we assessed the validity of a refined version of the report indifference points procedure, where we altered the instructions to try and make them easier to understand. This involved adding a description of what some example choices would mean. Secondly, we included the quiz that was used in Johnson et al. (2015). The precise wording of the quiz was obtained from Johnson via email. We assessed the validity of the refined report indifference points procedure in four different ways: 1) we examined levels of unsystematic discounting (as in Experiment 2A), 2) participants rated how much they thought they understood the instructions and how easy it was for them to answer the questions and we examined whether ratings were lower for participants with unsystematic data , 3) we assessed whether it can produce a well-established finding within the discounting literature (the magnitude effect; as in Johnson et al., 2015), and 4) we incorporated a second task that was designed to test the predictive validity of participants' self-reported indifference points (the *distance to indifference points procedure*). The distance to indifference points procedure allowed us to test the validity of Johnson and Bickel's (2008) criteria for detecting unsystematic discounting. If indifference points reported by the systematic participants have higher predictive validity than indifference points reported by the unsystematic participants, then this would indicate that Johnson and Bickel's criteria are a valid way of detecting discounting functions that do not accurately describe the value of the larger, delayed reward for an individual.

Thus, as a secondary outcome to testing the report indifference points procedure, Experiment 2B also allowed us to evaluate Johnson and Bickel's (2008) criteria for systematic discounting. Johnson and Bickel's criteria are widely used (Smith et al., 2018) but have not been directly evaluated. Using these criteria, the results of Experiment 2A suggest that the report indifference points procedure is not suitable for use in subsequent experiments because it produces an unacceptable number of unsystematic discounting functions. However, if the criteria are inaccurate, this conclusion may not be warranted.

Participants completed the report indifference points procedure first, and we then used the indifference points each participant reported to generate specific, individualized predictions that we tested in the distance to indifference points procedure. Our use of the distance to indifference points procedure to validate the report indifference points procedure is a novel contribution to the literature. In the distance to indifference points procedure, participants made a series of choices between a smaller, immediate reward and a larger reward received after various delays, as in the titrating-amount procedure. The delays and the size of the larger, delayed reward were the same in both the report indifference points and distance to indifference points procedures. However, the size of the smaller, sooner reward in the distance to indifference points procedure was specific to each participant, based on their indifference points obtained from the report indifference points procedure. That is, the smaller, immediate reward was the participant's previously reported indifference point multiplied by five different values ('distances'): 0.5, 0.75, 1, 1.25 and 1.5. Multiplying an indifference point by these values would produce smaller, immediate amounts of varying distances from the original indifference point. Thus, we will refer to these values as 'distances' throughout the thesis. Multiplying an indifference point by the distances that are less than 1 would produce smaller, immediate amounts that are lower in value than the participant's previously reported indifference point. Therefore, participants should choose the larger, delayed reward at distances less than 1 if their previously reported indifference points were accurate. However, multiplying an indifference point by the distances that are greater than 1 would produce smaller, immediate amounts that are higher in value than the participant's previously reported indifference point. Therefore, participants should choose the smaller, immediate reward at distances greater than 1. Participants should choose the smaller, immediate reward approximately 50% of the time at the distance of 1, indicating indifference. If we plot the distance to the indifference point against the number of times the smaller, immediate reward was chosen, then this particular pattern of results should have a sigmoid-like shape, as shown in Figure 4.3.

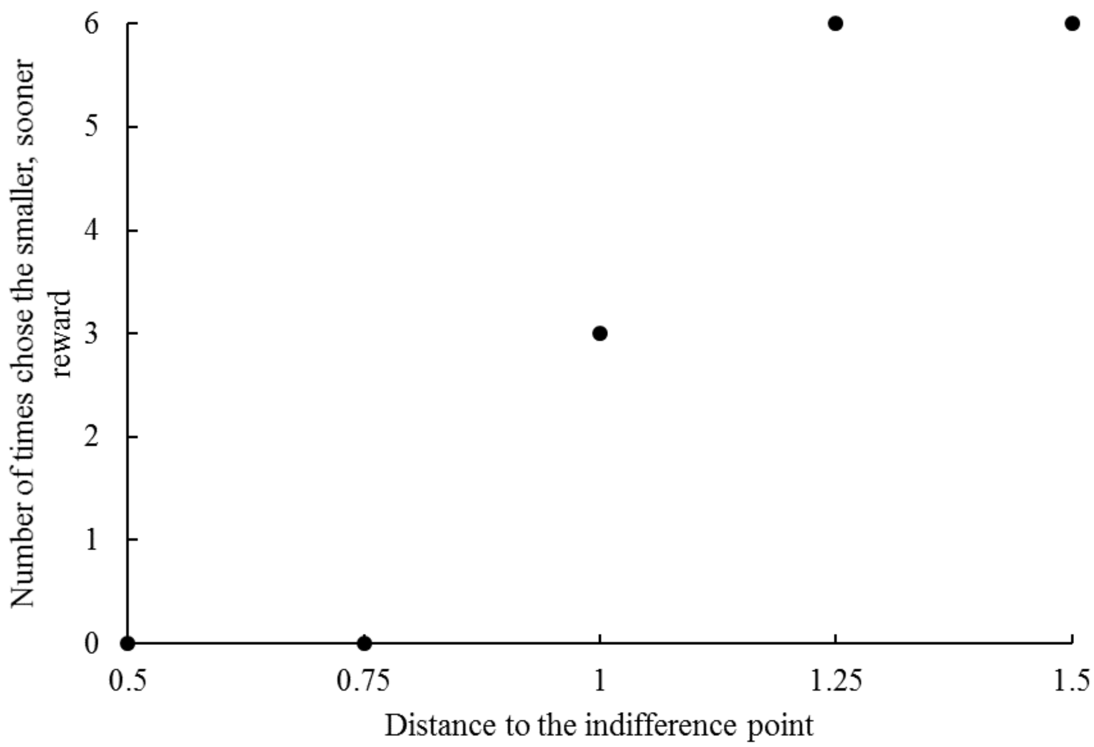

Figure 4.3. Example of what the data from the distance to indifference points procedure would look like if choices are consistent with choices made in the report indifference points procedure. The values on the y axis range from zero to six, given that there were six delays and therefore six choices in total for each distance in this example (consistent with the procedural details of Experiment 2B; see Method section).

This pattern of choices would indicate that the report indifference points procedure produces valid data, as it would show that participants made choices in the distance to indifference points procedure that are consistent with the choices made in the report indifference points procedure. We therefore fitted a logistic function (sigmoid curve) to these data in order to test the predictive validity of participants' reported indifference points. Equation 4.1 shows the logistic function (Verhulst, 1847; Richards, 1959):

$$y = \mathrm{L} + \frac{\mathrm{U} - \mathrm{L}}{1 + e^{-k(x - x_0)}} \qquad \text{[Equation 4.1]}$$

Where *y* is the y axis value (i.e., the number of times the smaller, sooner reward was chosen), *L* is the lower asymptote (i.e., the curve's minimum value), *U* is the upper asymptote (i.e., the curve's maximum value), *e* is the natural logarithm base (i.e., Euler's number), *k* is the steepness of the curve, *x* is the x axis value (i.e., the distance to the indifference point), and x_0 is the x axis value at the midpoint of the curve. *L*, *U*, *k* and x_0 were free-parameters, the value of which were identified by fitting the logistic function to the data points.

If the logistic function describes the data well, then this would suggest that participants made choices in the distance to indifference points procedure that were systematically related to their reported indifference points. However, there are two problems with interpreting R^2 values from models with multiple free-parameters: 1) Every time a parameter is added to a model, the R^2 increases, even if due to chance alone, and 2) if a model has too many parameters, then it may begin to model random noise within the data (this is referred to as *overfitting*; Burnham & Anderson, 2002). To correct for the first problem, we calculated adjusted R^2 which accounts for the number of free parameters, as shown in Equation 4.2:

$$R^2 = 1 - \frac{\mathrm{SSR}/(n-p)}{\mathrm{SST}/(n-1)} \qquad \text{[Equation 4.2]}$$

Where *SSR* is the sum of squares of the residuals from the curve, *n* is the number of data points, *p* is the number of free-parameters and *SST* is the total sum of squares. Although adjusted R^2 is not designed to detect overfitting, note that a substantial proportion of the unsystematic participants still had very low adjusted R^2s (see Results and Discussion), indicating that the model was not over fitted for those individuals. Additionally, the primary aim of this analysis was to compare adjusted R^2s between two different groups (participants who showed systematic discounting vs. unsystematic discounting in the report indifference points procedure), for which the same model was used with the same number of free parameters.

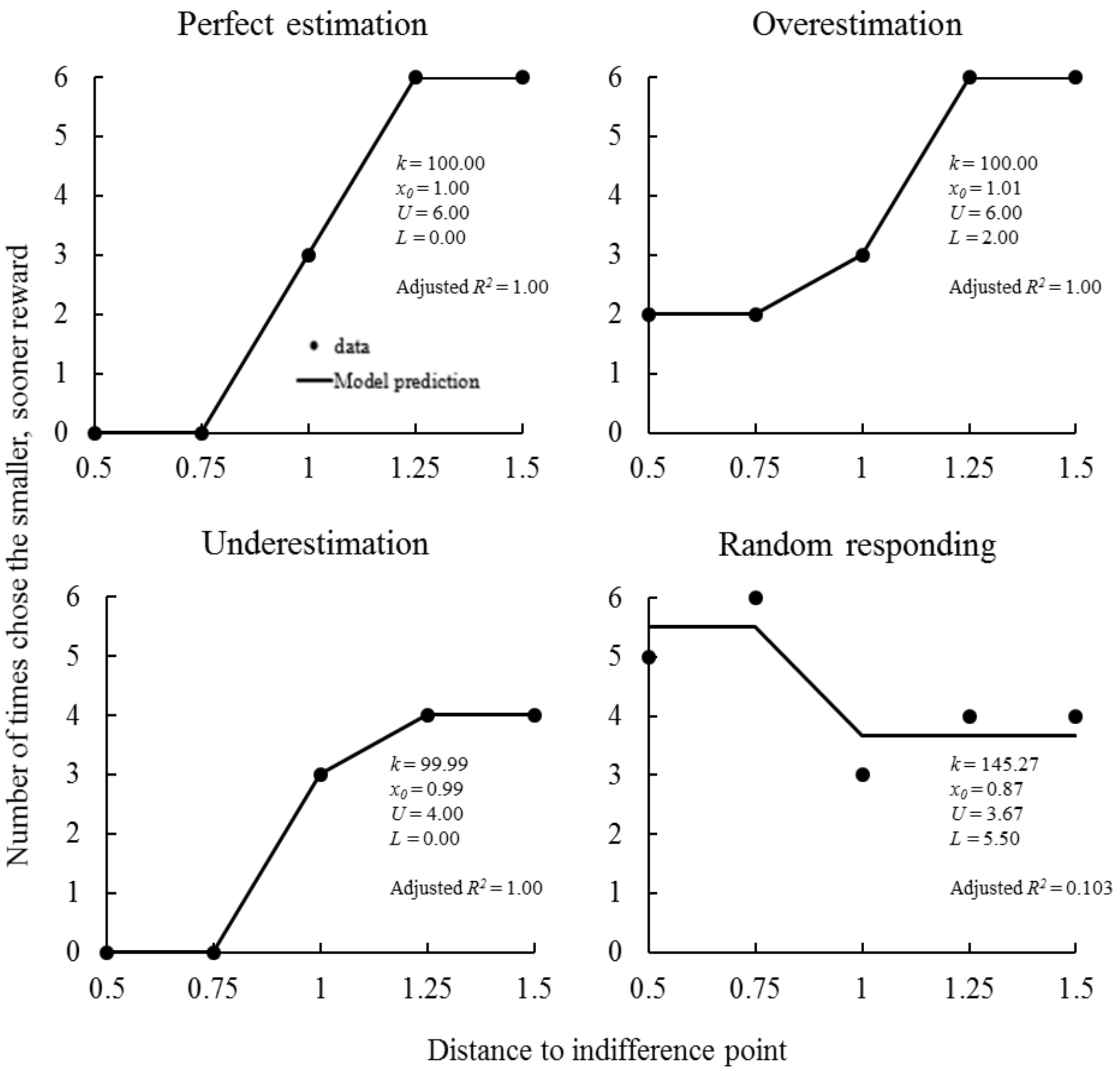

*Figure 4.4.*The logistic function (black line) fitted to hypothetical data points from the distance to indifference points procedure. The values of the free parameters (k, x_0, U and L) and model fits (adjusted R^2s) of the logistic function are also shown. The different graphs show examples of what the data and model fits would look like if: 1) choices are perfectly consistent with choices made in the report indifference points procedure (top left 'perfect estimation' graph), 2) participants overestimated their indifference points in the report indifference points procedure (top right 'overestimation' graph), 3) participants underestimated their indifference points in the report indifference points procedure (bottom left 'underestimation' graph), and 4) choices are inconsistent with choices made in the report indifference points procedure (bottom right 'random responding' graph).

Figure 4.4 shows the logistic function (black line) fitted to hypothetical data points for illustrative purposes. The top left graph of Figure 4.4 shows what the data and model fit would look like if an individual made choices in the distance to indifference points procedure that perfectly corresponded to their reported indifference points. The top right graph shows an example where an individual made choices in the distance to indifference points procedure that were systematically related to their reported indifference points (adjusted $R^2 = 1.00$), although also indicates that the individual *overestimated* their indifference points ($L > 0$). Put another way, say, for example, an individual reported in the report indifference points procedure that $700 now made them feel just as good as $1,000 in 1 year. However, during the distance to indifference points procedure, they chose $350 now over $1,000 in 1 year (on a 0.5 distance trial). This would indicate that the individual overestimated their indifference point in the report indifference points procedure, given that $350 is less than $700. The bottom left graph also shows an example where an individual made choices in the distance to indifference points procedure that were systematically related to their reported indifference points (adjusted $R^2 = 1.00$), although in this instance the indifference points were *underestimated* ($U < 6$). The bottom right graph shows an example where an individual made choices in the distance to indifference points procedure that were *not* systematically related to their reported indifference points (adjusted R^2 is low, indicating random responding).[11]

We compared the free-parameters and adjusted R^2 values between participants who previously showed systematic discounting and participants who previously showed unsystematic discounting in the report indifference points procedure. We expected the logistic function to describe systematic participants' data better than unsystematic participants' data. This result would show that Johnson and Bickel's (2008) criteria are a valid way of detecting discounting functions that do not accurately describe the value of the larger, delayed reward for an individual.

If participants who showed unsystematic discounting in the report indifference points procedure have higher *L* values (i.e., higher lower asymptotes) than participants who showed systematic discounting, then this would suggest that the unsystematic participants overestimated their indifference points in the report indifference points procedure relative to the systematic participants. If unsystematic participants have lower *U* values (i.e., lower upper asymptotes) than the systematic participants, then this would suggest that the unsystematic participants underestimated their indifference points in the report indifference

[11] A participant might also overestimate their indifference points to such a high degree that they choose the smaller, sooner reward at all delays in the distance to indifference points procedure. This particular pattern of responding would produce a very low adjusted R^2 which would be indicative of random responding, even though the choice patterns between the two procedures would be related. However, few participants from Experiments 2B and 2C showed this particular pattern of responding (see Appendix G).

points procedure relative to the systematic participants. The former result would be consistent with Experiment 2A, as when the unsystematic participants in Experiment 2A experienced the report indifference points procedure first (which is comparable to the current experiment), discounting was significantly shallower in the report indifference points procedure (*Mdn* AUC = 0.91) than in the titrating-amount procedure (*Mdn* = 0.60), $p < 0.001$. In contrast, when the systematic participants in Experiment 2A experienced the report indifference points procedure first, discounting was steeper in the report indifference points procedure than in the titrating-amount procedure, indicating that the systematic participants underestimated their indifference points.

In summary, Experiment 2B investigated the validity of a refined version of the report indifference points procedure. Evidence for the validity of the procedure would include: 1) low levels of unsystematic discounting, 2) participants reporting that they found the questions easy to understand, 3) a magnitude effect, and 4) predictions derived from the report indifference points procedure being confirmed by the distance to indifference points procedure. Additionally, if indifference points reported by the systematic participants have higher predictive validity than indifference points reported by the unsystematic participants, then this would support the validity of Johnson and Bickel's (2008) criteria for identifying unsystematic discounting.

Method

Participants

102 participants were recruited online via Mechanical Turk in exchange for $1.50 USD which was payed into their Amazon payment accounts upon completion. The survey was only available to master workers who had previously demonstrated high quality responses across tasks on Mechanical Turk. Amazon does not disclose how they designate master workers, but it is likely based on ratings from Mechanical Turk requesters that workers have previously completed tasks for. The mean age of the sample was 36.13 (*SD* = 9.17) and 60% were males. Informed consent was obtained prior to participation.

Materials

Participants completed the following tasks online using Qualtrics Survey Software (Qualtrics, Provo, UT).

Report indifference points procedure. Participants read the same instructions for this task as in Experiment 2A, except that they were told that they would be answering 12 questions, rather than six. Following these instructions, they were also given two examples to read (with a screenshot of what the slider tool would look like for each scenario). Figure 4.5 shows a screenshot of the examples.

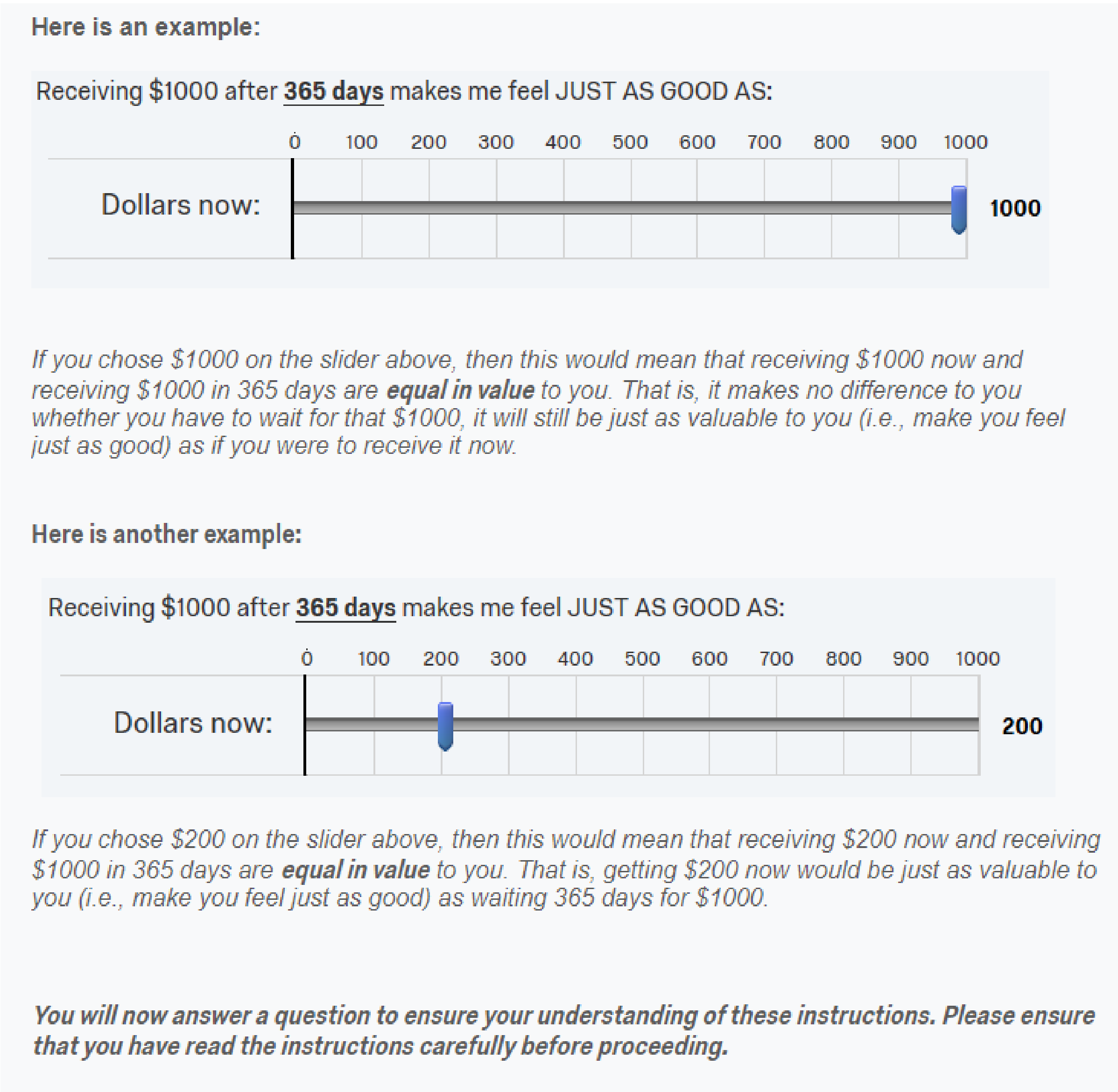

Figure 4.5. Screenshot of the examples participants read after reading the instructions for the report indifference points procedure.

Participants were then asked the following multiple-choice question (taken from Johnson et al., 2015): "*For the next twelve questions, you should use the slider tool to tell us:*" The options they could choose were:

1. how much later you would like to receive the money,
2. the amount of money you could choose to receive now that would make you feel better than the money you could choose to receive later,
3. the amount of money you would choose to receive now that would make you feel just as good as the money you could choose to receive later, or
4. the amount of money you could choose to receive now that would make you feel worse than the money you would choose to receive later.

Participants could not proceed to the next part of the task until they had correctly answered this question (i.e., selected option 3). If participants selected the incorrect option, then the following message appeared: *"Your answer is incorrect. Please try again."* The order of the multiple-choice options was randomised for every attempt.

Once participants had correctly answered the above question, they proceeded to the report indifference points task. The report indifference points task was identical to Experiment 2A, except that there were six questions where the larger, delayed amount was $100 (in addition to the six questions where the larger, delayed amount was $1,000). Secondly, all trials/choices were presented in a randomised order on the same screen, meaning that participants had the option to go back and view/change their previous choices, unlike in Experiment 2A. We randomised the questions to try and more closely replicate Johnson et al. (2015). However, it was not possible to present the questions on separate pages and also record reaction times for the distance to indifference points procedure (see below), which was a vital measure used for a different research project within our lab (for which the same participants' data were used; Stone, 2018).

Following the report indifference points procedure, participants were asked to rate how much they thought they understood the instructions using a slider tool which ranged from 0 to 100 (0 = did not understand the instructions at all, 50 = unsure, 100 = completely understood). They then rated how easy it was for them to answer the questions (0 = difficult, 50 = about average, 100 = easy).

Distance to indifference points procedure. Participants read the following instructions:

> Thank you!
>
> You have completed the first portion of the survey.
>
> If you need to take a break, stretch your legs, or move away from the computer, please do so now.
>
> For the next phase of the survey, you will be asked a series of similar hypothetical (pretend) questions.
>
> That is, you will be asked to choose between two alternatives: to receive some money now or to receive an alternative amount after a certain period of time. Your job is to click on the option that you would prefer.
>
> Please answer each question as quickly as you feel comfortable.
>
> Try to avoid taking any breaks or moving away from your computer screen.

Participants made a series of choices between a smaller, immediate reward that varied in size and a larger reward ($100 or $1,000) that varied in the delay to its receipt. There were six delays, identical to the delays used in the report indifference points procedure. The sizes of the smaller, immediate rewards were specific to each participant, based on their self-reported indifference points from the report indifference points procedure. That is, the smaller, immediate rewards were calculated as the self-reported indifference points multiplied by five different values (i.e., ‘distances’): 0.5, 0.75, 1, 1.25, or 1.5. For example, if a participant had reported an indifference point of $75 for $100 delayed by 14 days , then during the current task, that participant would be asked to choose between smaller, immediate amounts of $37.50, $56.25, $75, $93.75 $112.50 and the larger, delayed amount of $100 in 14 days. There were 60 trials/choices in total (i.e., amount [2] x delay [6] x distance [5]), presented in a randomised order.

Procedure

Participants completed the report indifference points procedure followed by the distance to indifference points procedure. Finally, they were presented with a written debrief on their computer screen which explained the purpose of the experiment. The entire survey took approximately 15 minutes to complete.

Data analyses

We fit the hyperbolic model to each individual's data from the report indifference points procedure for each larger, delayed amount condition. We also calculated the AUCs, as in Experiment 2A. Tests of normality (Shapiro Wilk and visual inspection of histograms depicting data distributions) showed that all AUC and *k* values were not normally distributed (all *p*s > 0.05), and log transformations failed to normalise these values. We therefore used a Wilcoxon-signed rank test to determine if discounting was significantly different between the two larger, delayed amount conditions. Given the high levels of unsystematic discounting (see Results and Discussion), results are reported when unsystematic participants were included as well as when they were excluded[12], as in Experiment 2A.

The comprehension and easiness ratings for the report indifference points procedure were not normally distributed (*p*s > 0.05). We therefore compared these ratings between the systematic and unsystematic participants using Mann-Whitney U tests.

We calculated the number of times the smaller, immediate reward was chosen for each individual at each distance to the indifference point for each larger, delayed amount in the distance to indifference points procedure. These values ranged from zero to six, given that there were six delays (and therefore six choices for each distance). We used Microsoft Excel Solver® to fit the logistic function (Equation 4.1) to each participant's data points for each larger, delayed amount, and identified the L, U, k, and x_0 values. All free-parameters were not normally distributed (all *p*s > 0.05). We therefore compared these values between systematic and unsystematic participants using Mann-Whitney U tests using a Bonferroni-corrected alpha level of 0.013, given the high number of comparisons. We also calculated adjusted R^2 values (using Equation 4.2, as previously described). The adjusted R^2 values were not normally distributed (*p*s < 0.05); we therefore used Mann-Whitney U tests to compare the adjusted R^2s between the systematic and unsystematic participants.

Results and Discussion

Levels of unsystematic discounting

A total of 57% of participants had unsystematic discounting in either one or both monetary conditions from the report indifference points procedure. Table 4.2 shows the

[12] If a participant had unsystematic discounting in one or both magnitude conditions, then that participant's entire dataset was excluded. That is, they did not contribute to the group median for either magnitude.

percentage of participants with unsystematic discounting based on each of Johnson and Bickel's (2008) criteria. These percentages are well above the typical range and are again similar to Experiment 2A and Rung et al. (2018).

Table 4.2

Percentage of participants with unsystematic discounting (based on Johnson & Bickel's 2008 criteria) in the $100 and $1,000 conditions.

Condition	Criterion 1 only	Criterion 2 only	Both criteria	Total
$100	7%	24% (flat function = 7%)	14%	44%
$1,000	4%	28% (flat function = 8%)	11%	43%

Magnitude effect

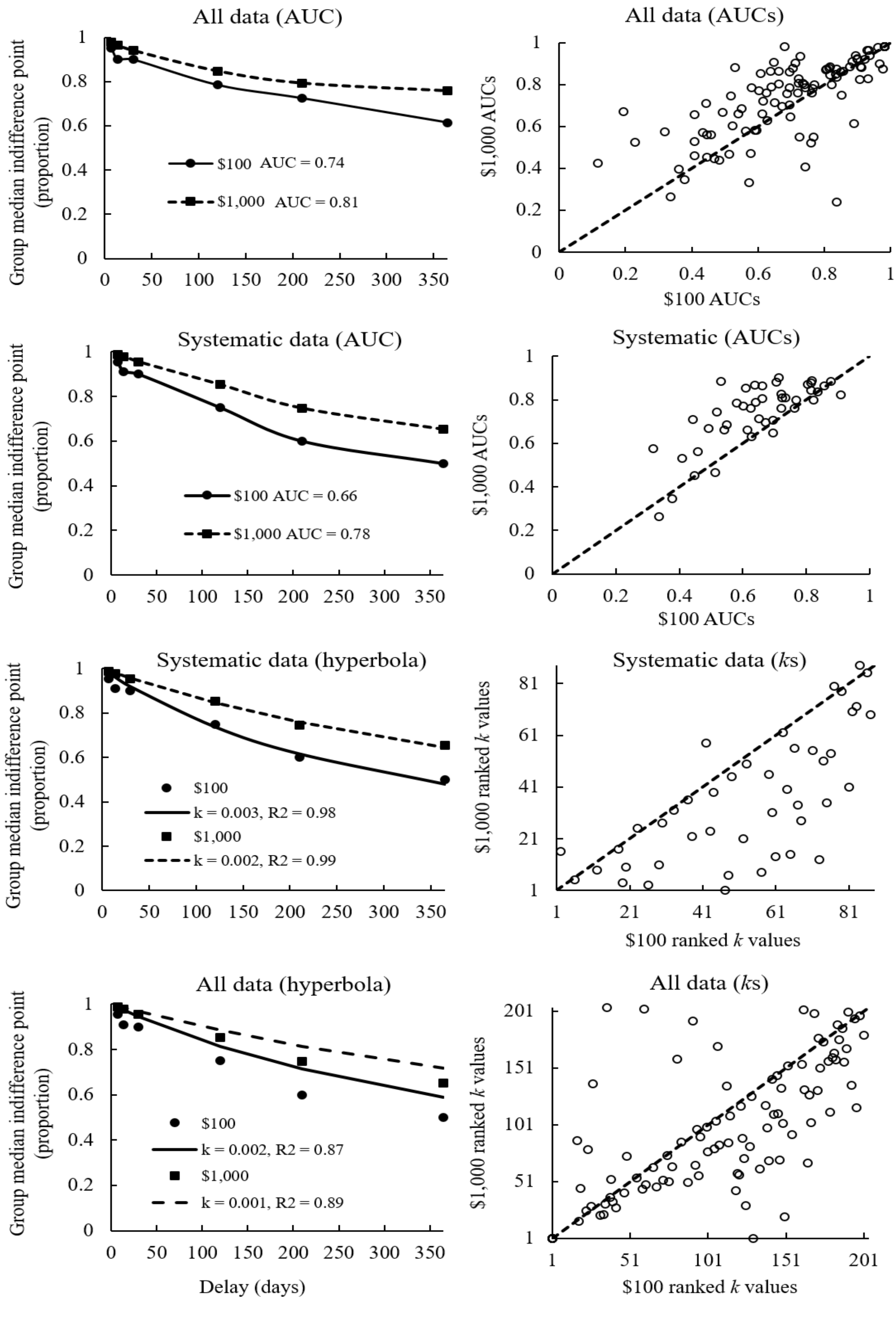

Figure 4.6. The left panel graphs depict the group median data. The top left graph shows the group median indifference points (calculated from all participants' data expressed as a proportion of the larger, delayed amount) plotted against the delays for the $100 condition

(circles and solid line) and the $1,000 condition (squares and dashed line) from the report indifference points procedure. The data points are joined by the lines to visually depict the area under the discounting curve (AUC). The second left graph is the same as the top left graph except that only systematic participants are included. In the third left graph the lines depict the hyperbolic model fitted to the group median indifference points (calculated from systematic participants' data only). The bottom left graph is the same as the third left graph except that all participants are included. The right panel graphs depict the individual data. The top right graph shows the AUCs for the $100 condition plotted against the AUCs for the $1,000 condition for all individuals, whereas the second right graph shows these data for systematic individuals only. The third right graph shows the *k* values for the $100 condition plotted against the *k* values for the $1,000 condition for the systematic individuals only, whereas the bottom right graph shows these data for all individuals. The ranks of the *k* values are plotted so that all values are visible despite the skewed distribution. Within each of the right panel graphs, one circle represents one individual. Circles above the dashed line represent participants with higher AUC or *k* values in the $1,000 condition, whereas circles below the dashed line represent participants with higher AUC or *k* values in the $100 condition. Note that lower AUCs and higher *k*s indicate steeper discounting.

Participants showed significantly steeper discounting in the $100 condition than in the $1,000 condition according to both AUC and *k* values, regardless of whether unsystematic participants were excluded (all ps < 0.001). This pattern was evident at both the group (left panel of Figure 4.6) and individual (right panel of Figure 4.6) level. As shown in the top two right-panel graphs, most participants had higher AUCs in the $1,000 condition than in the $100 condition, regardless of whether unsystematic participants were excluded (all participants = 60%; systematic participants = 70%). As shown in the bottom two right-panel graphs, most participants had higher *k* values in the $100 condition than in the $1,000 condition, regardless of whether unsystematic participants were excluded (all participants = 62%; systematic participants = 89%). These results show a clear magnitude effect, consistent with Johnson et al. (2015). Additionally, discounting rates in the $100 condition were significantly correlated with discounting rates in the $1,000 condition (all ps < 0.001). In contrast to the weak correlations found between the report indifference points procedure and the titrating-amount procedure in Experiment 2A, these correlations were strong (r_s = 0.70-0.75), confirming that the two conditions measured similar constructs.

Although the group median indifference points were well described by the hyperbolic model (as shown in Figure 4.6), discounting was not well described by the hyperbolic model at the individual level ($100 *Mdn* R^2 = 0.36; $1,000 *Mdn* R^2 = 0.61), as in Experiment 2A. Additionally, although there were no systematic trends in the residuals (see Appendix E), almost one quarter (23%) of the systematic participants still had data in one or both monetary conditions that were not well-described by the hyperbolic model (R^2s < 0.5). While this percentage is reduced compared to Experiment 2A (one third vs. one quarter), the hyperbolic model did not describe the data well for a substantial proportion of these individuals.

Comprehension of the report indifference points procedure

Overall, participants reported that they understood the instructions of the report indifference points procedure (group median rating = 100). The group minimum comprehension rating was 60, however removal of one outlier increased the group minimum to 71. Participants also rated the questions as easy to answer (group median rating = 90.5). The group minimum easiness rating was 8, however, only 11% of participants gave ratings of 50 or less (i.e., rated the questions as average or below average). There was no significant difference in comprehension ratings between the systematic and unsystematic participants (both *Mdn*s = 100). There was also no significant difference in easiness ratings between the two groups (systematic *Mdn* = 92; unsystematic group *Mdn* = 90; U = 1161.00, p = 0.432).

Distance to indifference points procedure

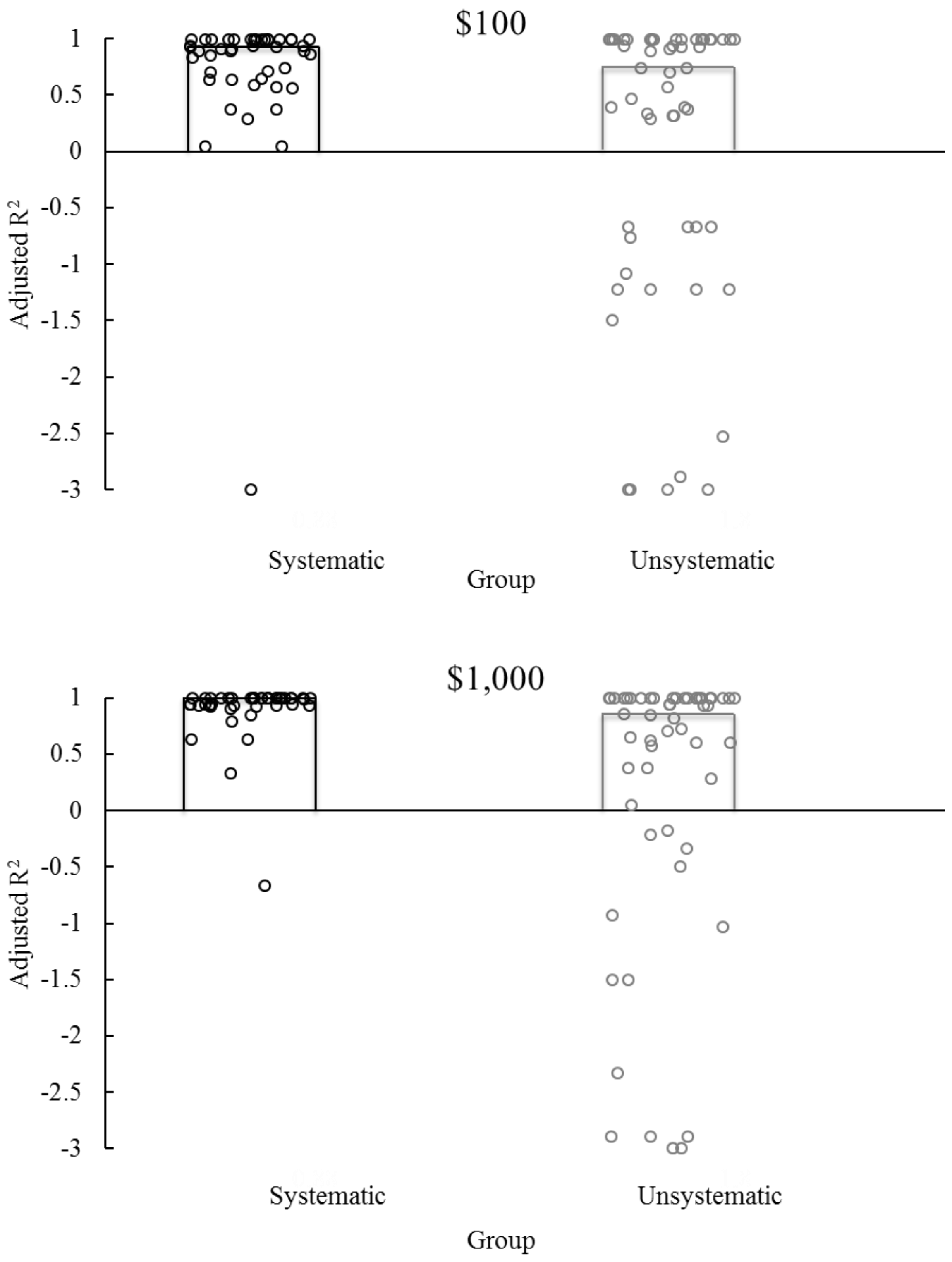

Figure 4.7. Logistic model adjusted R^2s from the distance to indifference points procedure for participants who had systematic discounting (left black bars and circles) and for participants who had unsystematic discounting (right grey bars and circles) in the report indifference points procedure. The top graph shows the adjusted R^2s for the $100 condition and the bottom graph shows the adjusted R^2s for the $1,000 condition. The bars represent the group median adjusted R^2s and the circles represent the individual adjusted R^2s. The circles are jittered so that overlapping circles are visible.

Figure 4.7 shows the group median and individual logistic function adjusted R^2s from the distance to indifference points procedure for participants who had systematic discounting and for participants who had unsystematic discounting in the report indifference points procedure. The unsystematic participants had significantly lower R^2s than the systematic participants for both larger, delayed amounts ($100: Systematic *Mdn* = 0.94; unsystematic *Mdn* = 0.75; U = 931.50, p = 0.049; $1,000: Systematic *Mdn* = 1.00; unsystematic *Mdn* = 0.85; U = 820.00, p = 0.008). As shown in Figure 4.7 the logistic model did not describe the data well for almost half of the unsystematic participants (38% and 48% of participants had R^2s lower than 0.5 in the $1,000 and $100 conditions, respectively). Conversely, only 5% and 14% of the systematic participants had R^2s lower than 0.5 in the $1,000 and $100 conditions, respectively. Most systematic participants had R^2s greater than 0.7 in the $1,000 (91%) and $100 (73%) conditions. The finding that the logistic model described the systematic participants' data better than the unsystematic participants' data indicates that the unsystematic participants made choices in the distance to indifference points procedure that were less consistent with their previously reported indifference points. This supports the validity of Johnson and Bickel's (2008) criteria for identifying unsystematic discounting.

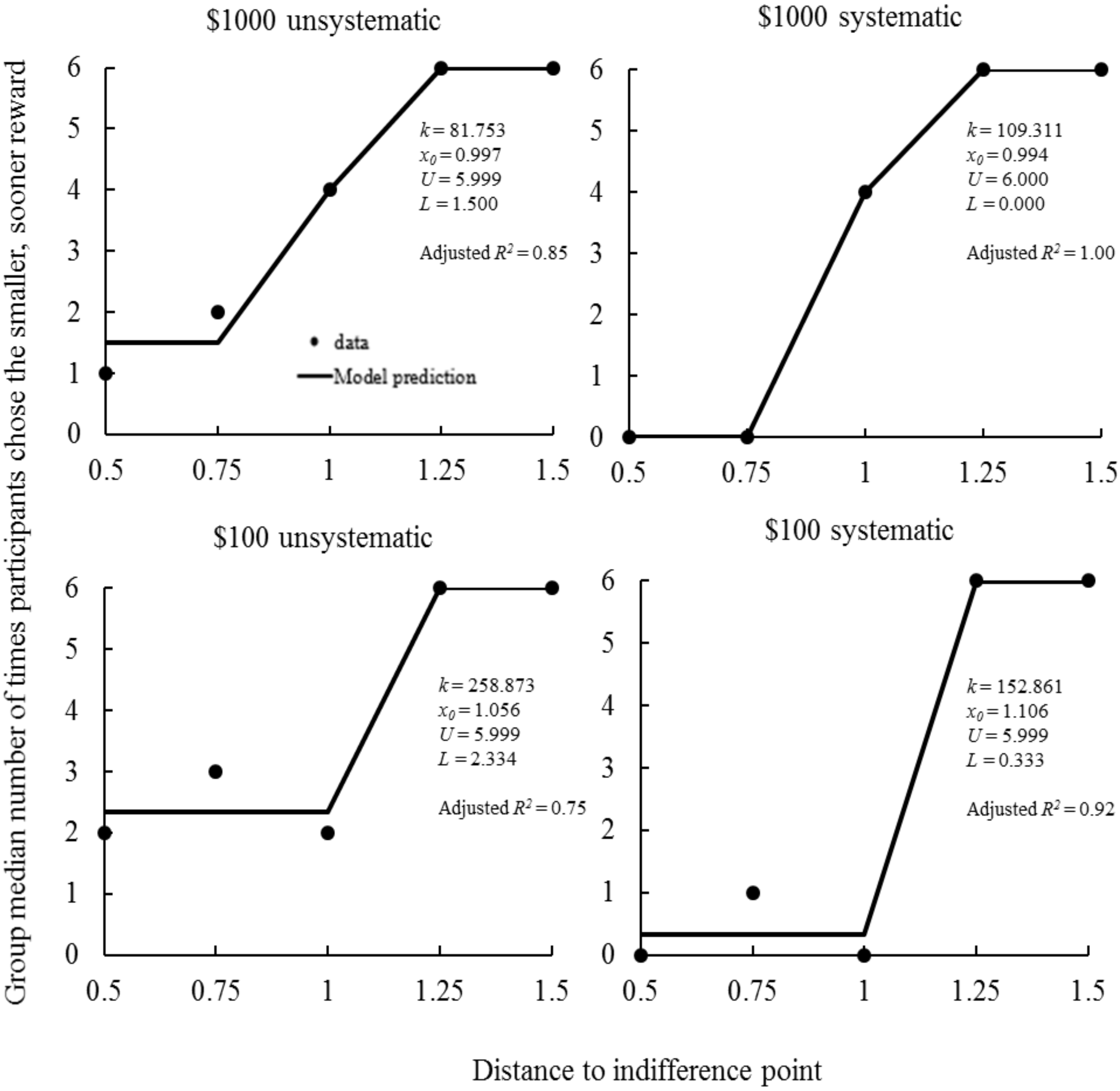

Figure 4.8. The logistic function (black line) fitted to the group median number of times participants chose the smaller, sooner reward (black points) at each distance to the indifference point. The top graphs represent participants in the $1,000 condition who previously showed unsystematic (left graph) and systematic (right graph) discounting in the report indifference points procedure. The bottom graphs represent participants in the $100 condition who previously showed unsystematic (left graph) and systematic (right graph) discounting in the report indifference points procedure. The values of the free parameters (*k*, x_0, *U* and *L*) and model fits (adjusted R^2s) of the logistic function are also shown on each graph.

Figure 4.8 shows the logistic function fitted to the group median number of times participants chose the smaller, sooner reward at each distance to the indifference point. The top graphs show participants in the $1,000 condition who previously showed unsystematic and systematic discounting in the report indifference points procedure, and the bottom graphs show these data for the $100 condition. The unsystematic participants had higher *L* values (i.e., higher lower asymptotes) than the systematic participants for both monetary conditions, however, this difference was only significant for the $1,000 condition. There were no significant differences in *k*, x_0 or *U* values between the systematic and unsystematic groups, and these values approximated those expected if choices in the distance to indifference point procedure were well predicted by the report indifference points procedure (see Figure 4.4, top left graph). Table 4.3 shows the medians and Mann-Whitney *p* values for each comparison.

Table 4.3

Group median logistic model L, k, x_0 and U values from the distance to indifference points procedure for participants who showed systematic and unsystematic discounting in the report indifference points procedure. The p values were obtained from Mann-Whitney U tests.

$100			
Parameter	Systematic	Unsystematic	*p*-value
L	0.42	1.83	0.079
k	100.00	100.00	N/A
x_0	1.10	1.10	N/A
U	6.00	5.65	0.270
$1,000			
Parameter	Systematic	Unsystematic	*p*-value
L	0.00	0.50	0.004*
k	99.59	100.00	0.419
x_0	0.99	1.00	0.495
U	6.00	6.00	N/A

Note: N/A indicates that a statistical test was unnecessary because the group medians were identical.
* indicates that the *p*-value was statistically significant at the Bonferroni-corrected alpha level of 0.01.

The finding that the *L* values were significantly higher in the unsystematic group in the $1,000 condition is consistent with the explanation that the unsystematic participants overestimated their indifference points relative to the systematic participants. However, although the logistic function described the group median data well (as shown in Figure 4.8), this was not the case at the individual level, as a substantial proportion of unsystematic participants did not have data that were well-described by the logistic function, as previously shown in Figure 4.7 (see Appendix G for figures depicting the logistic function fitted to the data for each unsystematic individual where adjusted $R^2 < 0.5$). Thus, we also directly compared the number of smaller, sooner choices between the groups at each distance to the indifference point. This allowed us to determine whether the relative overestimation by the unsystematic group was also evident when using an analysis that was not limited by a poor model fit. Consistent with the results using the parameters of the logistic function, unsystematic participants made significantly more smaller, sooner choices at the 0.5 (U = 901.50, p = 0.004) and 0.75 (U = 893.00, p = 0.006) distances compared to the systematic group in the $1,000 condition. There were no significant differences between the two groups at the 1, 1.25 and 1.5 distances (as the medians were identical, as shown in Figure 4.8). In contrast to the results using the logistic function parameters, unsystematic participants made significantly more smaller, sooner choices at the 0.5 (U = 901.50, p = 0.003), 0.75 (U = 893.00, p = 0.004) and 1 (U = 893.00, p = 0.003) distances compared to the systematic group in the $100 condition. There were no significant differences between the two groups at the 1.25 and 1.5 distances (as the medians were the same; see Figure 4.8). This pattern of results in the $100 condition is consistent with the results in the $1,000 condition, in that the unsystematic participants overestimated their indifference points relative to the systematic participants. These findings are also consistent with Experiment 2A.

Conclusion

Although the refined version of the report indifference points procedure produced the well-established magnitude effect, it also produced unacceptable levels of unsystematic discounting, indicating that this procedure will not be appropriate to use for future studies, consistent with Experiment 2A. Both the systematic and unsystematic participants reported that they had a good understanding of the instructions overall, and that they found the questions easy to answer. However, the finding that a substantial proportion of the unsystematic participants made choices in the distance to indifference points procedure that were not well described by the logistic function, and thus not systematically related to their reported indifference points, suggests that these participants did not accurately report their own indifference points. This pattern of responding was less frequent for the systematic participants, indicating that Johnson and Bickel's (2008) criteria are a valid way of detecting discounting functions that do not accurately describe the value of the larger, delayed reward for an individual.

Experiment 2C

The reason for the different levels of unsystematic discounting between Johnson et al. (2015) and Experiment 2B is unclear. Although, one methodological difference that remains between the two studies is whether the questions were presented on the same screen. In Johnson et al.'s study, each choice was presented on its own page, meaning that participants clicked a 'next' button to move onto the next choice with no option to go back and view their previous choices. It is possible that having all choices presented at once in the current experiment increased participants' cognitive load which could explain the higher levels of unsystematic discounting (Franco-Watkins et al., 2010). Although it is perhaps unlikely that presenting all delays simultaneously increased cognitive load to an extent likely to cause the higher levels of unsystematic discounting we observed, it does remain a difference between the two procedures. We therefore conducted Experiment 2C to more closely replicate Johnson et al.'s report indifference points procedure.

Method

The method for Experiment 2C was identical to Experiment 2B, except that each trial/choice in the 'report indifference points' task was presented on its own page, as in Johnson et al. (2015). We also only included delayed amounts of $1,000, given that a clear magnitude effect had already been established in Experiment 2B. Participants were 52 master workers recruited online via Mechanical Turk in exchange for $1.50 USD. The mean age of the sample was 33.17 (*SD* = 6.94) and 75% were males. Informed consent was obtained prior to participation.

Results and Discussion

Levels of unsystematic discounting

A total of 37% of participants had unsystematic discounting according to Johnson and Bickel's (2008) criteria in the report indifference points procedure. Specifically, 10% had unsystematic discounting according to criterion 1 only, 19% had unsystematic discounting according to criterion 2 only (flat function = 6%), and 8% had unsystematic discounting according to both criteria 1 and 2. Although the overall level of unsystematic discounting is slightly lower than in single conditions from Experiment 2B (37% vs. 43-44%), it is still well above the typical range (18%; Smith et al., 2018).

Comprehension of the report indifference points procedure

Overall, participants reported that they understood the instructions of the report indifference points procedure (group median rating = 100). Only 2% of the participants' comprehension ratings were below 70. Participants also rated the questions as easy to answer (group median rating = 94.5). Only 8% of the participants gave easiness ratings of 50 or less (i.e., rated the questions as average or below average). The systematic group had significantly higher comprehension ratings (*Mdn* = 100) than the unsystematic group (*Mdn* = 99), U = 199.00, p = 0.008, although ratings in both groups were very high. There was no significant difference in easiness ratings between the two groups (systematic *Mdn* = 96; unsystematic group *Mdn* = 93; U = 284.00, p = 0.562). These results are consistent with Experiment 2B.

Distance to indifference points task

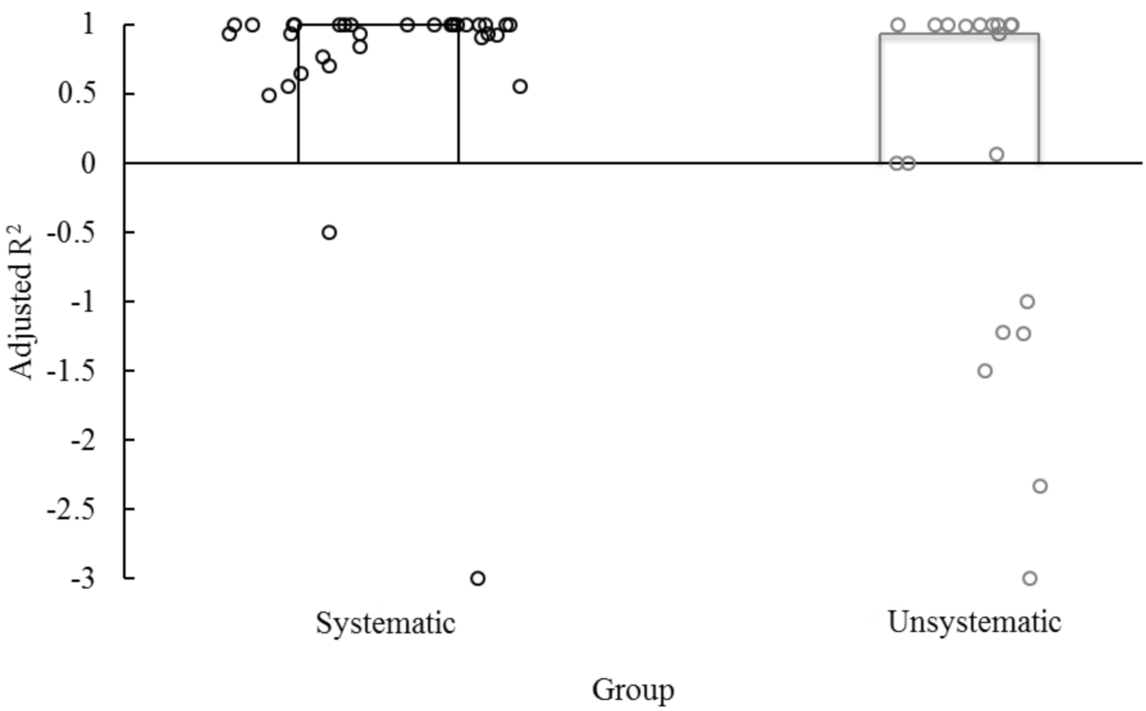

Figure 4.9. Logistic model adjusted R^2s from the distance to indifference points procedure for participants who had systematic discounting (left black bars and circles) and for participants who had unsystematic discounting (right grey bars and circles) in the report indifference points procedure. The bars represent the group median R^2s and the circles represent the individual R^2s. The circles are jittered so that overlapping circles are visible.

There was no significant difference in adjusted R^2 values between the systematic (*Mdn* = 0.999) and unsystematic (*Mdn* = 0.935) groups, U = 240.00, p = 0.160. However, as shown in Figure 4.9, the logistic model did not describe the data well for a substantial proportion of the unsystematic participants (47% had R^2 values below 0.5), as in Experiment 2B. Conversely, only 9% of the systematic participants had R^2s less than 0.5; most (82%) had

R^2s above 0.7. Thus, the individual data indicate that the logistic model described the systematic participants' data better than the unsystematic participants' data. The non-significant result from the Mann-Whitney U test is likely due to the bimodal distribution of the data for the unsystematic participants. As can be seen in Figure 4.9, approximately half of the unsystematic participants had very high R^2s (greater than 0.9), whereas the other half had very low R^2s (less than 0.1), which would have produced a high group median R^2.

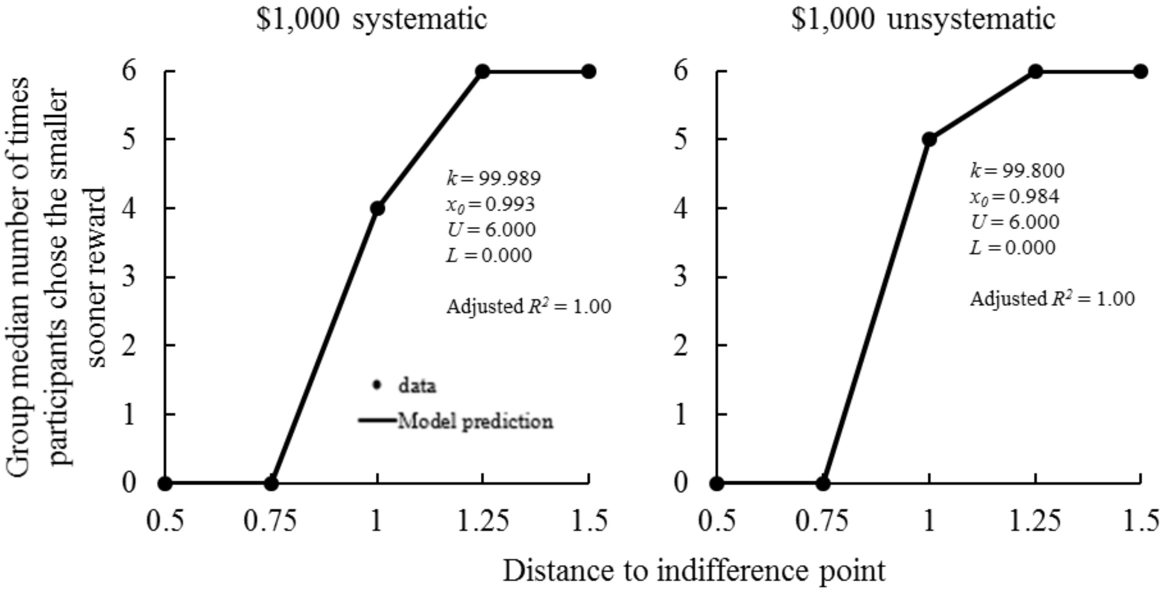

Figure 4.10. The logistic function (black line) fitted to the group median number of times participants chose the smaller, sooner reward (black points) at each distance to the indifference point. The left graph represents participants who previously showed systematic discounting and the right graph represents participants who previously showed unsystematic discounting in the report indifference points task. The values of the free parameters (k, x_0, U and L) and model fits (adjusted R^2s) of the logistic function are also shown on each graph.

Figure 4.10 shows the logistic function fitted to the group median number of times participants chose the smaller, sooner reward at each distance to the indifference point (see Appendix G for figures depicting the logistic function fitted to data from each unsystematic individual where adjusted $R^2 < 0.5$). There was no significant difference in k (unsystematic Mdn = 100, systematic Mdn = 99.69, U = 259.50, p = 0.304), x_0 (unsystematic Mdn = 0.902, systematic Mdn = 0.993, U = 268.00, p = 0.387), U (unsystematic Mdn = 6.00, systematic Mdn = 5.75, U = 299.00, p = 0.783) or L (both Mdns = 0.00) values between the systematic and unsystematic groups. Consistent with the results using the parameters of the logistic function, there was no significant difference in the number of times participants chose the smaller, sooner reward at any distance to the indifference point (distance of 1: U = 269.00, p

= 0.386; all other medians were the same; see Figure 4.10). Thus, in contrast to Experiment 2B, there was no evidence to suggest that the unsystematic participants overestimated their indifference points.

This inconsistency between Experiments 2B and 2C may be due to the fact that we did not categorise whether participants were unsystematic for each larger delayed amount separately in Experiment 2B, meaning that some of the 'unsystematic' participants may have only had unsystematic indifference points in one of the magnitude conditions. When we reanalysed the data with participants categorised as unsystematic or systematic for each amount condition separately, Experiment 2B results were consistent with Experiment 2A results. That is, there was no significant difference in *L* values (or in the number of times participants chose the smaller, sooner reward) between the unsystematic and systematic groups for the $1,000 condition (all ps > 0.05). The unsystematic participants still had significantly higher *L* values (and a larger number of smaller, sooner choices at the 0.5 and 0.75 distances) than the systematic participants in the $100 condition, however. The percentage of systematic participants with low adjusted R^2s (< 0.5) also increased (from 14% to 19% in the $100 condition and from 5% to 16% in the $1,000 condition) when we re-categorised the systematic/unsystematic participants separately for each delayed amount. This finding that more of the systematic participants had poor logistic function R^2s in Experiment 2B than in the current experiment (16-19% vs. 9%) suggests that participants may be more likely to report inaccurate indifference points when they are required to make more choices (as in Experiment 2B, which would produce an even higher cognitive load).

Conclusion

Findings from Experiment 2C were consistent with Experiment 2B, as the report indifference points procedure produced unacceptable levels of unsystematic discounting. Furthermore, almost half of the unsystematic participants made choices in the distance to indifference points procedure that were not systematically related to their reported indifference points, indicating that they did not accurately report their own indifference points. This pattern of responding was less frequent for the systematic participants, which again supports the validity of Johnson and Bickel's (2008) criteria for detecting unsystematic discounting. In conclusion, the report indifference points procedure produces a high percentage of invalid data, a finding that was replicated across three different experiments, two different types of participants (university students and Mechanical Turk participants) and two different larger, delayed amounts ($100 and $1,000). We therefore continued to use the well-established titrating-amount procedure in subsequent studies of the current thesis.

Chapter 5: Investigating the necessary components of EFT to reduce delay discounting

Chapter 5 details two experiments where the primary aim was to determine the components of EFT that are necessary to reduce delay discounting. Experiment 3A investigated the necessity and relative importance of the episodic and future thinking components of EFT, and Experiment 3B investigated whether the future events need to be personally relevant to reduce discounting.

Experiment 3A

Studies have shown that thinking about the future is more effective at reducing delay discounting than thinking about the past (e.g. Dassen et al., 2016. O'Donnell et al., 2017, Stein et al., 2017). These studies, however, do not tell us whether thinking about the future is necessary to reduce delay discounting rate or whether thinking about the past also has a significant but smaller effect. To test this question, studies need to compare discounting rates between an Episodic Past Thinking (EPT) and a No Thinking (NT) control condition. If EPT reduces discounting rate compared to a NT control, then this would show that thinking about the future is not a necessary component of the EFT effect. To reiterate, the temporal attention hypothesis (Radu et al., 2011) asserts that delay discounting rate may be reduced by shifting attention away from present, which suggests that thinking about the past may be sufficient to reduce delay discounting rate. The Constructive Episodic Simulation (CES) hypothesis (Schacter & Addis, 2007) also suggests that EPT may have a similar effect to EFT on discounting rate, as it posits that people use memories to construct an imagined future. Functional magnetic imaging studies indicate that similar brain regions, such as the hippocampus, are used when individuals remember the past and imagine the future (Addis, Wong, & Schacter, 2007; Botzung, Denkova, & Manning, 2008).

The question of whether EPT and EFT have interchangeable effects on delay discounting rate has been tested by Daniel et al. (2016). Participants were randomly allocated to an EFT condition (where participants thought about events up to 6 months into the future), an EPT condition (where participants thought about events from up to 6 months ago from the past), or an Episodic Recent Thinking (ERT) control condition. In the ERT condition, participants thought about positive events that happened within the last 24 hours and also positive events that will happen within the next 24 hours. The researchers tested the effects of both EFT and EPT on delay discounting of future outcomes (using choices between a smaller reward now vs. a larger, later reward) and on delay discounting of past outcomes (using choices between a smaller reward now vs. a larger reward in the past). An example of delay discounting of past outcomes would be choosing to eat cake now even though you achieved weight loss in the past. Thus, the health rewards of the past outcome are discounted in favour of the immediate gratification of unhealthy food. Results showed that EFT reduced delay discounting of future outcomes, but not delay discounting of past outcomes. Additionally, EPT reduced delay discounting of past outcomes, but not delay discounting of future outcomes. The authors concluded that, although thinking about the future and thinking about the past may draw on similar brain systems, EFT and EPT effects on delay discounting are not interchangeable. However, the control condition used in this study was an ERT condition, rather than a NT control condition. Thus, Daniel et al.'s study does not rule out the possibility that both EPT and ERT reduce discounting (although not as much as EFT) compared to no manipulation.

To date, only one study has compared the effect of EPT on delay discounting rate to a NT control condition (Lempert, Speer, Delgado, & Phelps, 2017). In this within-subjects design study, participants wrote about positive, personally relevant memories involving a specific time and place. Three days later they completed the hypothetical money task where they were cued to think about a different memory every six choices. During the control NT block, participants were cued to "relax" every 6 choices. Participants had lower discounting rates in the EPT blocks compared to the control blocks. Lempert et al.'s (2017) findings suggest that future thinking is not necessary to reduce delay discounting rate. It is therefore possible that Daniel et al.'s (2016) EPT condition might have reduced discounting of future outcomes compared to baseline but this was not detected because the ERT condition also had an effect.

Additionally, studies have shown that episodic thinking is more effective at reducing delay discounting than semantic thinking (e.g. Benoit et al., 2011; Chiou & Wu, 2017). These studies, however, do not tell us whether episodic thinking is necessary to reduce delay discounting rate. In order to test this question, studies would need to compare discounting rates between a Semantic Future Thinking (SFT) and a NT control condition. If SFT reduces discounting rate compared to a NT control, then this would show that episodic thinking is not a necessary component of the EFT effect. Episodic thinking may not be a necessary component of the EFT effect, as EFT has been shown to reduce delay discounting rate (compared to a NT control condition) in individuals with impaired episodic prospection abilities (i.e., patients with medial-temporal lobe damage; Kwan et al., 2015). Kwan et al. (2015) speculated that their results might be explained by personally relevant, SFT processes within these individuals. For example, if participants were cued to imagine attending a friend's wedding (EFT), they could have instead imagined what tends to happen at a wedding, based on their semantic knowledge. Conversely, Palombo et al. (2015) found no effect of EFT on discounting rate in individuals with similar impairments to those in Kwan et al.'s study. As discussed by Kwan et al., a notable difference between the two studies was that participants in Kwan et al.'s study generated their own, personally relevant future events, whereas Palombo et al. provided participants with events without ensuring personal relevance. Personally-/emotionally-relevant events are associated with larger EFT effects (Benoit et al., 2011) which could explain the difference in findings of the two studies.

To date, few studies have compared SFT to a NT condition in order to further address the necessity of episodic thinking in reducing delay discounting rate (Chiou & Wu, 2017; Palombo, Keane, & Verfaellie, 2016; Wu et al., 2017). In Palombo et al.'s (2016) study, healthy participants completed a NT control condition (a standard hypothetical money task), followed by a SFT condition. In the SFT condition, participants were asked to generate an item that they could purchase in the future (e.g., *"what could you buy for $42 in 4 months?"*) before they made each choice between a smaller, immediate reward and a larger, delayed reward (e.g., $30 now or $42 in 4 months). Participants had lower discounting rates in the SFT condition compared to the NT control condition. Additionally, participants stated that they mostly thought about single objects (not a scene) in the SFT condition. Although it is possible that episodic thinking was employed to some extent during the SFT condition, this manipulation did not *require* episodic thinking. Thus, these results suggest that episodic thinking may not be necessary to reduce discounting rate, consistent with Kwan et al. (2015).

Chiou and Wu (2017) and Wu et al. (2017) also compared the effects of SFT and NT conditions on delay discounting rate, however their SFT manipulation largely differed from that in Palombo et al.'s (2016) study. In Chiou and Wu's study, nicotine-dependent individuals either: 1) wrote about three events that could happen in one year if they quit smoking now (EFT condition), 2) wrote about three positive benefits that would occur in one year if they quit smoking now (SFT condition), or 3) completed an unrelated questionnaire (NT control condition). All groups then completed a delay discounting task. Participants in the EFT condition had lower discounting rates than participants in the NT condition.

However, in the contrast to Palombo et al., there was no significant difference in discounting between the SFT and NT conditions.

In Wu et al.'s (2017) study, university students either: 1) wrote about three aspects of their ideal selves (e.g., physical, social, moral and psychological) and then listed three events that would occur "if the desirable aspects of the self are realized in 1 year's time." (EFT condition), 2) wrote about aspects of their ideal selves only (that would occur in 1 year; SFT), or 3) completed an unrelated questionnaire (NT condition). Consistent with Chiou and Wu (2017), participants in the EFT condition had lower discounting rates than participants in the NT condition, but there was no significant difference in discounting between the SFT and NT conditions.

The inconsistent results between Palombo et al. (2016) and Chiou and Wu (2017) and Wu et al. (2017) may be explained by the different ways that SFT was manipulated. In Palombo et al.'s study, the SFT manipulation was particularly relevant to the delay discounting task in that both directly involved money. Although it is possible that participants considered financial implications during Chiou and Wu's and Wu et al.'s SFT manipulations (such as the financial cost of being their ideal physical selves), this is less likely to be the case compared to in Palombo et al.'s SFT manipulation.

In summary, little research to date has directly assessed whether the future and episodic components of EFT are necessary to reduce delay discounting. Only one study has investigated EPT in comparison to a NT control, while studies investigating the necessity of episodic thinking have produced inconsistent results. Additionally, no studies have assessed the relative importance of both the episodic and future thinking components within the same experiment. The current study sought to assess this by comparing the effects of EFT, EPT, SFT and NT (control) conditions on delay discounting rate. In the SFT condition in the current study, participants were instructed to identify items that they could purchase with $1,000 in the future. We manipulated SFT similarly to Benoit et al. (2011) and Palombo et al. (2016), rather than as in Chiou and Wu (2017) and Wu et al. (2017), because this SFT manipulation is more comparable to EFT in that a similar degree of content from the manipulation overlaps with content from the money task. That is, the future delays in the EFT condition matched the future delays in the discounting task, and the amount of money in the SFT condition matched the larger, delayed amount in the discounting task. Although an overlap in content between the SFT and EFT tasks and the money task could inflate demand characteristics (in contrast to the EPT condition), this was tested for in the current study (see below). In contrast to Palombo et al., we did not prompt participants to identify items that they could purchase at specific delays, reasoning that localising this in time could encourage episodic thinking (i.e., visualising purchasing the item at a specific time with a scene unfolding). Although Palombo et al. found that participants mostly thought about only objects in their SFT condition, there were still some instances in which participants reported thinking about a scene unfolding; perhaps this was enough episodic thinking to reduce discounting rate. We also added additional instructions telling participants what to think about when presented with the semantic cues during the money task (what they generally know about the item) to further discourage episodic thinking.

We predicted that the EPT condition would reduce discounting compared to the NT control condition, based on the temporal attention hypothesis, the CES hypothesis and Lempert et al. (2017). We also predicted that the SFT condition would reduce discounting compared to the NT control condition, based on results from Kwan et al. (2015) and Palombo et al. (2016). We expected the EFT condition to produce the lowest discounting rates, consistent with previous research (e.g., Dassen et al., 2016; O'Donnell et al., 2017; Stein et al., 2017; Benoit et al., 2011; Chiou & Wu, 2017; Wu et al., 2017). These set of results would suggest that, while episodic and future thinking are *most effective* at reducing delay discounting rate, they are not *necessary* to reduce delay discounting rate. We were also interested in assessing the *relative* importance of the episodic and future thinking components. For example, if discounting rate is lower in the EPT condition than in the SFT condition then this would suggest that episodic thinking is the more important component, rather than future thinking.

We also sought to determine whether EFT (or EPT or SFT) has a lasting effect on discounting rate over time, or whether discounting rate is only decreased while cues are present. This was achieved by presenting participants with the NT condition twice; once at the beginning of the experiment (control time 1) and once again at the end of the experiment (control time 2). The experimental conditions were presented in a random order between the two NT control conditions. If participants have lower discounting rates in the control time 2 condition than in the control time 1 condition, then this would suggest that EFT has a lasting effect on delay discounting rate. Finally, we included a set of questions that assessed what participants thought the research question and hypotheses were for the current experiment (based on Stein et al., 2018), to determine whether demand characteristics could explain any of our results.

Method

Participants

Participants were 96 first-year psychology students at Victoria University of Wellington recruited in partial fulfilment of a course requirement. No other demographic information was collected. Informed consent was obtained prior to participation.

Materials

Participants completed all of the following tasks in the laboratory on computers. The tasks were constructed using Microsoft Visual Basic® 2017 Software**.**

Control 'no thinking' condition. Participants read the following instructions:

> You will now be asked a set of hypothetical questions that require you to make choices between different amounts of money. Please read each question carefully as the amount of money you can receive and when you can receive the money will vary. There is no right or wrong answer, select the option you would be most likely to choose.
>
> The word 'choose' will be presented during these scenarios, simply prompting you to answer the question by making your choice.

The participants were asked to indicate their preference between two hypothetical choices: one reward that was smaller in magnitude and available immediately ($500 now) and a larger reward that was available after a delay (receive $1,000 after X delay). There were six delays, one randomly generated for each participant from each of the following ranges: 1-7 days, about 2 weeks (11-17 days), about 1 month (21-35 days), 3-4 months (84-126 days), 6-8 months (168-245 days), 1 year (336-391 days). The delays were presented in a random order. The word 'choose' was written in red centred above the two choices. An indifference point was identified for each delay using the titrating amount procedure, as previously described. The mouse cursor returned to the space between the two alternatives after each trial so that it was equally effortful to choose either alternative.

After completing the money task, participants were asked: *What were you thinking about when making choices between the different amounts of money?* Participants typed their answer.

EFT condition. Participants were asked to write about six vivid, emotionally positive, personally relevant and exciting future events (one for each delay range; 0-7 days, about 2 weeks, about 1 month, 3-4 months, 6-8 months, 1 year). They were then asked to select on a calendar when the event will or could occur, and to type a cue word or phrase for each event. The calendar dates participants could select were constrained within the delay ranges as specified above for the control condition. The exact instructions given were identical to the instructions given in Experiment 1.

Next participants completed the hypothetical money task which was identical to the control condition except that: 1) the six delays used in the money task matched the delays that participants had previously selected on the calendar, 2) the cue word or phrase for the event corresponding to each delay was written in red centred above the two choices, rather than the word 'choose', and 3) participants read the following instructions regarding the cues:

> The cue words you generated for your imagined events will be presented during these scenarios. You do not have to base your choice on the imagined event; YOU JUST HAVE TO VIVIDLY IMAGINE THAT THIS EVENT IS REALLY HAPPENING, and next you can choose your preferred reward.

EPT condition. The EPT condition was identical to the EFT condition except that participants were asked to write about six vivid, emotionally positive, personally relevant and exciting events that happened in the *past* (one for each delay range; 0-7 days *ago,* about 2 weeks *ago*, about a month *ago*, about 3-4 months *ago*, about 6-8 months *ago*, about 1 year *ago*). The exact instructions were identical to the EFT condition except that a different tense was used so that they were asked to remember a past event, rather than imagine a future event. For example, if a participant had previously generated the cue phrase "birthday party" for their remembered event from 7 days ago, then during the money task "birthday party" would be written in red centred above the two choices when the delay was 7 days.

SFT condition. Participants read the following instructions:

> We want you to identify six, EQUALLY POSITIVE AND EXCITING items that a person could buy with approximately $1,000 in the future. An example

> could be a piece of furniture, an electronic device, or a piece of sports equipment.
>
> These descriptions will be used in the next exercise. Describe each item below. There is also a box for you to simply enter what the item is.
>
> Type your description of one item that a person could buy with $1,000 in the box below. Write what you generally know about the item (e.g., what does it look like? Where could a person buy the item from?).

After writing about the six items, participants completed the hypothetical money task where an item that they had previously written about was written in red centred above the two choices. The six delays used in the money task were randomly generated from each range as in the control condition. Participants read the following instructions regarding the cues:

> The items that you wrote about in the previous exercise will be presented during these scenarios. You do not have to base your choice on that item; YOU JUST HAVE TO THINK ABOUT WHAT YOU GENERALLY KNOW ABOUT THAT ITEM, and next you can choose your preferred reward.

All other instructions/details of the money task were the same as in the other conditions.

Manipulation check questions. As in Experiment 1, participants rated how often each cue made them think about the associated imagined event or item during the money task (1 = never, 7 = always), how vivid (1 = not vivid at all, 7 = highly vivid), positive (1 = not positive at all, 7 = highly positive), and exciting (1 = not exciting at all, 7 = highly exciting) the imagined event or item was, and how expensive the imagined event or item would be (1 = not expensive at all (very cheap), 7 = highly expensive).

Additionally, for each cue participants were asked: *What did you picture in your mind?* (as in Palombo et al., 2016). There were three options to choose from: 1) Nothing (i.e., either nothing or only a vague image was pictured), 2) Objects but not a scene (i.e., individual items or objects were pictured in isolation but not as part of a scene), or 3) A scene/scenario (i.e., an entire layout was pictured, including objects; the image could be static or dynamic, such as an unfolding scenario). Participants could only select one of the three options for each cue.

Questions to assess demand characteristics. Participants typed an answer to two open-ended questions:

1. Often, when people participate in a study like you are, they start to think about what the researcher is trying to study. In your own words, please tell us briefly what you think the purpose of this study is:
   The purpose of this study is to see…
2. Often, when people participate in a study like you are, they also start to think about what the researcher is expecting to find in their results. In your own words, please tell us briefly what you think the researcher's predictions are:
   I think the researcher is expecting to find…

The wording of Question 1 was taken from Stein et al. (2018). We included both of these questions to encourage participants to be specific about the researcher's hypothesis, as Question 2 asks a clearly different (and more specific) question from Question 1. That is, for Question 1, a participant might say "the purpose of this study is to see if thinking about events/items affects decisions about money," however, this answer does not tell us if the participant was aware of the direction in which we were expecting the events/items to affect their decisions (hence Question 2).

Procedure

All participants experienced the control condition first ('control time 1'), followed by the experimental conditions (i.e., EFT, EPT, SFT) in a random order, then finally the control condition again ('control time 2'). The manipulation check questions were presented at the end of each experimental condition. The entire experiment took two hours to complete. These two hours were broken up into two (one hour) sessions, separated by one week. In the first session, participants experienced the 'control time 1' condition and two experimental conditions. In the second session, participants experienced one experimental condition, then the 'control time 2' condition, then the questions to assess demand characteristics. Finally, participants were given a written and verbal debrief which explained the purpose of the experiment.

In order to link session 1 and session 2 data, participants were asked to provide a unique identifier for themselves (their birth date, birth month and current street name) at the end of session 1. At the start of session 2, participants were asked to select their unique identifier from a list, and were reminded that this was their birth date, birth month and current street name.

Data analyses

The hyperbolic model (Equation 1.1) was fit to each participants' indifference points for each discounting task condition, as in Experiment 1. We also calculated AUC (Equation 1.2) for each condition for each participant.

Tests of normality (Shapiro Wilk and visual inspection of histograms depicting data distributions) showed that k and AUC values were not normally distributed (all ps < 0.05, except for AUCs in the EPT condition where $p = 0.083$), and log transformations failed to normalise the data. Thus, a Friedman's test was used to determine if there was a significant main effect of condition. Post hoc comparisons were made using Wilcoxon signed-rank tests, with Bonferroni corrections reducing alpha level to 0.005.

For each participant, we calculated the number of times they stated that they pictured objects, a scene or nothing for each condition. These numbers were then converted into percentages by dividing the numbers by 6 and multiplying by 100. Tests of normality showed that answers to all manipulation check questions were not normally distributed (all ps < 0.05). We used Wilcoxon signed-rank tests to compare the percent of instances where participants stated that they pictured objects or a scene for each condition (Bonferroni-corrected alpha level = 0.007). For each of the other cue ratings (how exciting, expensive and positive each event/item was and the imagery score for each event/item), we conducted Friedman tests to assess whether there was a main effect of condition. Significant Friedman tests were followed up with post hoc comparisons using Wilcoxon signed-rank tests (Bonferroni-corrected alpha level = 0.017). We also compared the event ratings between the indifference points with the largest EFT effect and the indifference points with the smallest EFT effect using Wilcoxon signed-rank tests. The size of the EFT effect at each delay was determined by calculating the log ratio of the EFT indifference point and the control indifference point (log [EFT indifference point/control time 1 indifference point]).

For each participant, answers to both of the questions assessing demand characteristics were collated into one answer and coded into one of six categories: 1) guessed that the researcher was interested in how people make decisions about money, with no mention of the cues, 2) incorrectly guessed the researcher's prediction and results, 3) stated that they did not know the researcher's prediction, 4) correctly guessed the researcher's prediction, but this was inconsistent with the group results, 5) guessed the researcher's prediction which was consistent with the group results, or 6) correctly identified the research question, but did not clarify the direction of the prediction (e.g., "the purpose of this study is to assesses whether thinking about future events will affect decisions about money"). If a participant indicated that the researcher expected EFT, EPT and SFT to reduce discounting rate, then this would be a correct prediction, but would also need to match the results in order to be consistent with the presence of demand characteristics. Thus, only category 5 would be consistent with a demand characteristics interpretation. Alternatively, category 5 could indicate that participants were consciously aware of the decisions that they made during the money task, and therefore simply reported what they did, without this necessarily meaning that they changed their responses to conform to the researcher's predictions. However, an absence of (or few) participants in category 5 would rule out demand characteristics as an explanation for the pattern of results. The first five participants completed a version of the experiment without the questions assessing demand characteristics. See 'data exclusion' section below for further information on the final analysed sample size.

Unsystematic discounting and data exclusion

To reiterate, unsystematic data are usually identified and excluded if: 1) Any indifference point was larger than the previous indifference point by more than 20% of the larger, delayed reward, and 2) The last indifference point was not less than the first indifference point by at least 10% of the larger, delayed reward (Johnson & Bickel, 2008). As in Experiment 1, a large percentage of participants (39%) had unsystematic data in the EFT condition, according to Johnson and Bickel's (2008) criteria. The percentage of participants with unsystematic data in the other conditions was within the typical range (Smith et al., 2018). In the EPT and SFT conditions, 17% and 18% of participants had unsystematic discounting, respectively. In the control time 1 and control time 2 conditions, 5% and 10% of participants had unsystematic data, respectively. Table 5.1 shows the percentage of participants with unsystematic discounting according to each criteria in each condition.

Table 5.1
Percentage of participants with unsystematic discounting (based on Johnson & Bickel's 2008 criteria) in the EFT, EPT, SFT, control 1 and control 2 conditions.

Condition	Criterion 1 only	Criterion 2 only	Both criteria
Control 1	1%	4% (flat function = 4%)	0%
EFT	17%	14% (flat function = 3%)	8%
EPT	11%	3% (flat function = 3%)	3%
SFT	14%	3% (flat function = 2%)	1%
Control 2	3%	7% (flat function = 3%)	0%

Note: Flat function = the participants chose the larger, delayed reward on every trial.

The unsystematic discounting in the EFT condition is likely a consequence of the EFT manipulation, given that the different events were not matched in how positive, exciting, expensive or vivid they were. For example, a participant might show systematic discounting in the EFT condition until the final delay, where the indifference point greatly increases because the event at that particular delay is the most expensive or exciting. To test this explanation, we compared the excitingness, positivity, expensiveness and imagery score ratings between the EFT indifference points that positively deviated the most from the hyperbolic model and the EFT indifference points that deviated the least from the hyperbolic model using Wilcoxon signed rank tests. We determined the size of the difference between the observed EFT indifference points and the hyperbola-predicted EFT indifference points by calculating the log ratio (log [EFT indifference point/hyperbola-predicted EFT indifference point]), as in Experiment 1. These analyses are shown in the Results and Discussion section below.

Given that the unsystematic discounting in the experimental conditions may be a consequence of the manipulations, we only excluded participants with unsystematic data in control time 1 according to Johnson and Bickel's (2008) criteria (n = 4) for AUC analyses. We did not exclude participants based on these criteria for control time 2 in the reported results below, given that the experimental manipulations may have had a lasting effect. However, results did not change when these participants were excluded (*n* = 6). An additional four participants were excluded for not following instructions (e.g., they wrote about negative, rather than positive, events)[13]. Thus, the final analysed sample size was 87, except for the questions assessing demand characteristics where the final analysed sample size was 82.

Results and Discussion

K values did not produce any statistically significant results, unlike AUC results (see below). Although the group median *k*s trended in the same direction as the AUC results, the majority of the *k* comparisons were not significant at the strict Bonferroni-corrected alpha level of 0.005 (see Appendix H). This difference in results between AUC and *k* is consistent with the meta-analysis of the current thesis, which showed that EFT effect sizes were significantly smaller when *k* was used as a discounting rate measure. Additionally, inspection of the R^2 values revealed that discounting in the EFT condition was poorly described by the hyperbolic model (*Mdn* R^2s: Control 1 = 0.89; EFT = 0.59; EPT = 0.77; SFT = 0.83; control 2 = 0.85), as in Experiment 1. This is not surprising given that most of these individuals had unsystematic discounting according to criterion 1, and the hyperbolic model assumes that indifference points decrease as the delay to the larger amount increases (with the rate of decrease being faster over initial increases in delay). The model also poorly described discounting in the EFT condition for one quarter (25%) of the individuals who did not meet unsystematic criteria. The percentage of participants with these low R^2 values in the other conditions ranged from 10% to 14%. Examination of the residuals (indifference points predicted by the model minus observed indifference points) also showed that the model overestimated value at the shorter delays and underestimated value at the longest delay in the EFT condition (see Appendix I). Thus, AUC, rather than *k*, was used as the primary discounting rate measure.

Discounting rates (AUC)

[13] These four participants are also excluded from Table 5.1.

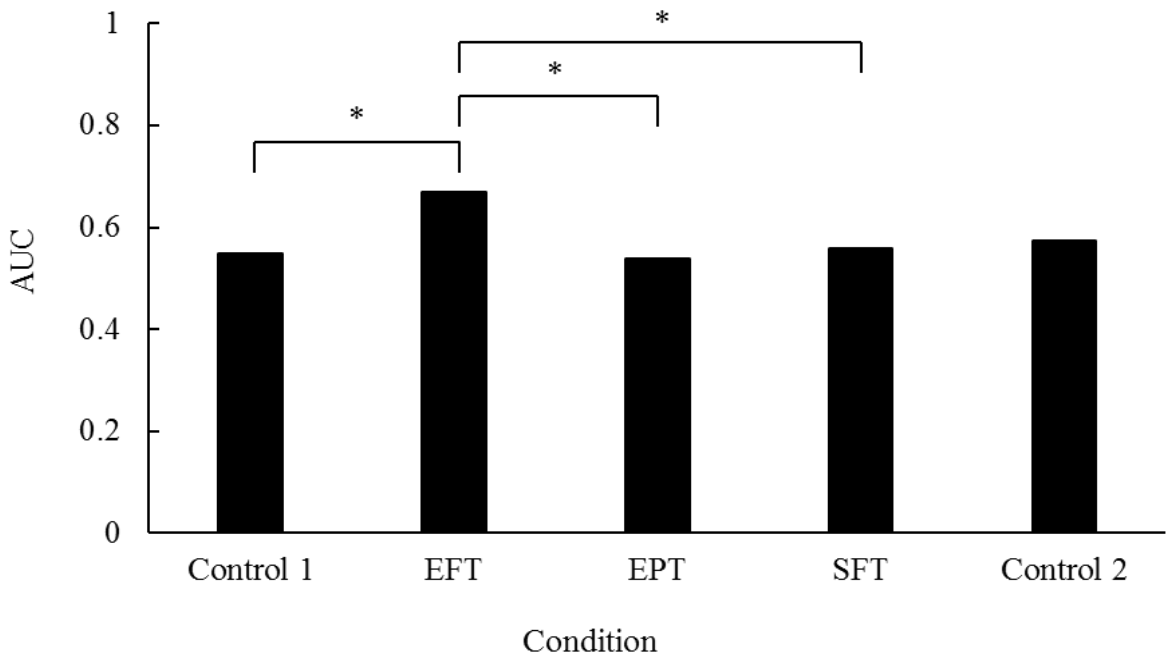

Figure 5.1. The group median AUC for the control time 1, EFT, EPT, SFT and control time 2 conditions. * indicates a significant difference between the two bracketed conditions.

Figure 5.1 shows the group median AUCs for each condition. A Friedman test revealed a significant main effect of condition, $X^2(4) = 27.29$, $p < 0.001$. Follow up Wilcoxon signed-rank tests showed that AUCs were significantly higher in the EFT condition compared to the control time 1 ($Z = -3.86$, $p < 0.001$), EPT ($Z = -3.61$ $p < 0.001$) and SFT ($Z = -3.72$, $p < 0.001$) conditions. As shown in Table 5.2, all other comparisons were non-significant.

Table 5.2
p values from post hoc comparisons using Wilcoxon signed-rank tests

	EFT	EPT	SFT	Control 2
Control 1	0.000*	0.454	0.896	0.011
EFT		0.000*	0.000*	0.009
EPT			0.902	0.068
SFT				0.113

* significant at Bonferroni-corrected alpha level of 0.005

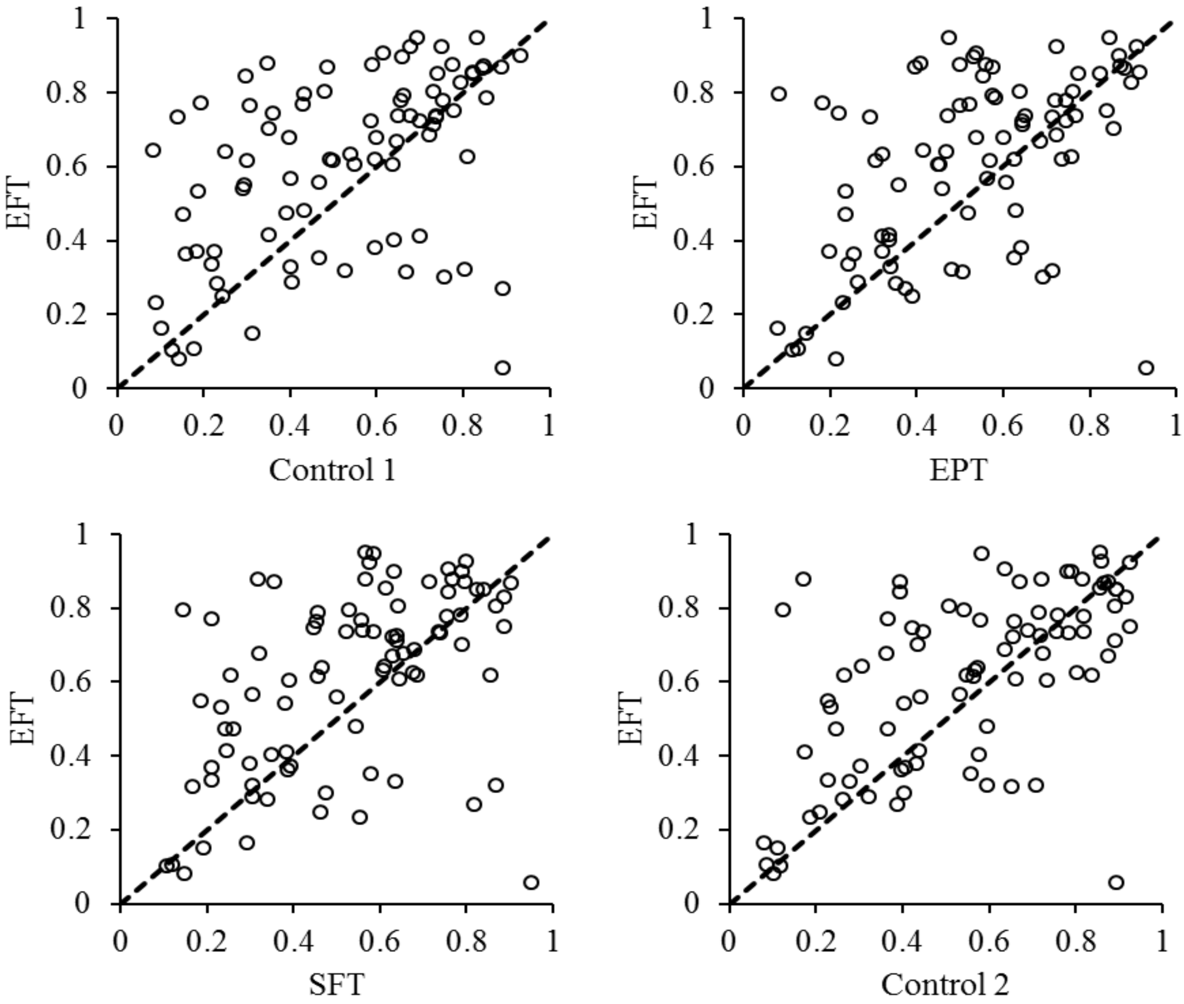

Figure 5.2. AUC values for the EFT condition plotted against the control time 1 (top left graph), EPT (top right graph), SFT (bottom left graph) and control time 2 (bottom right graph) conditions. For each graph, each point represents one participant. Plotted points above the dashed line represent participants with higher AUCs in the condition labelled on the y-axis (i.e., EFT). Plotted points below the dashed line represent participants with higher AUCs in the condition labelled on the x-axis.

Figure 5.2 depicts the individual data. The comparisons shown address the question of whether we replicated the general EFT effect. As shown in the top left graph of Figure 5.2, most participants (69%) had higher AUCs in the EFT condition compared to the control time 1 condition, and 3% of participants had similar AUCs in each condition (i.e., had points that overlapped with the reference line on the graph). As shown in the top right graph of Figure 5.2, most participants (62%) had higher AUCs in the EFT condition compared to the EPT condition, and 8% had similar AUCs in each condition. The bottom left graph shows that most participants (64%) had higher AUCs in the EFT condition compared to the SFT condition, and 6% had similar AUCs in each condition.

The bottom right graph of Figure 5.2 shows the AUCs for the EFT condition compared to the AUCs for the second control condition. Relative to the top left graph the number of participants that had higher AUCs in the EFT condition compared to the second control condition was reduced (57% vs 69%). The difference between EFT and control time 2 was not statistically significant at the Bonferroni-corrected alpha level of 0.005 (see Table 5.2).

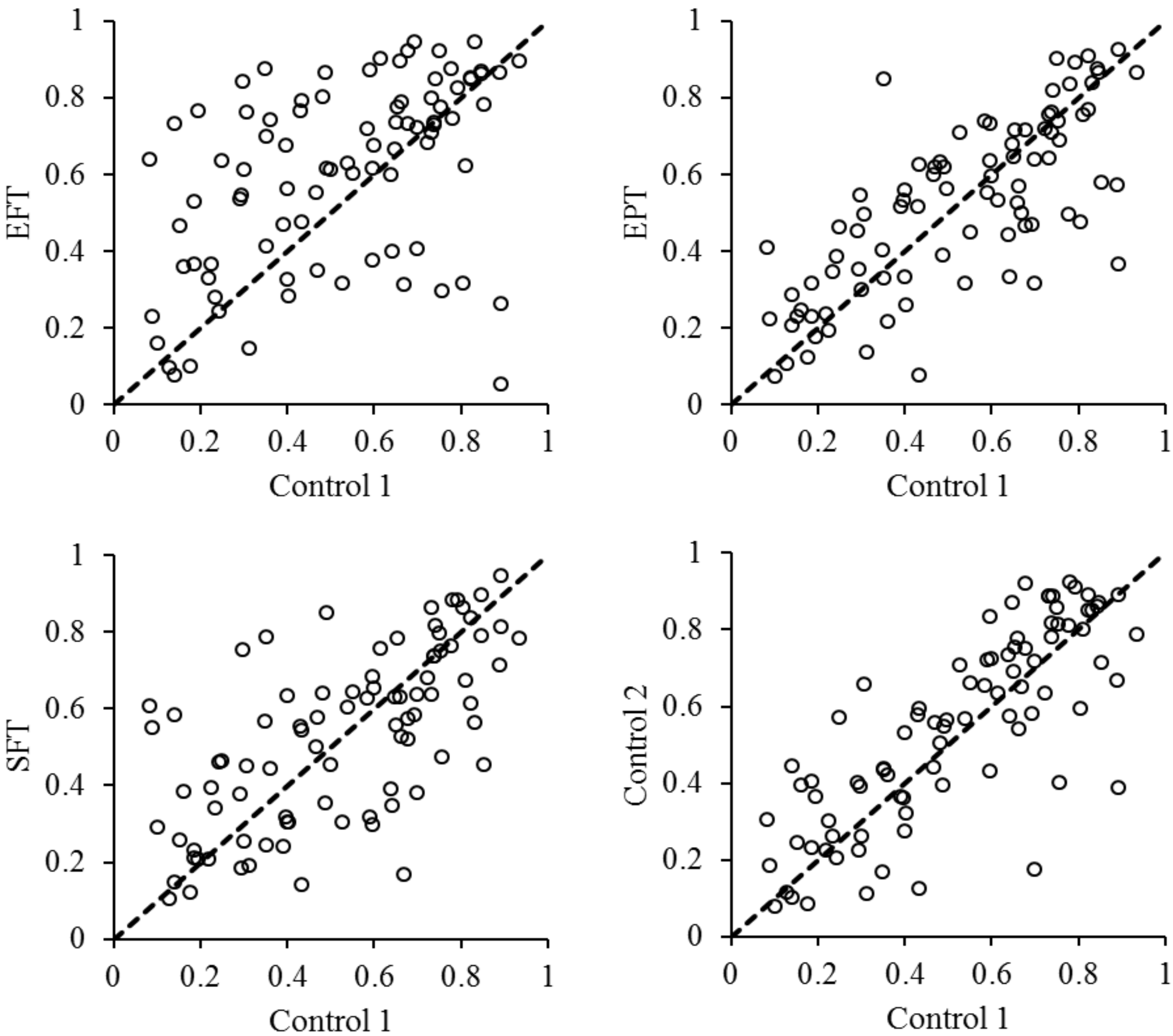

Figure 5.3. AUC values for the control time 1 condition plotted against the EFT (top left graph), EPT (top right graph), SFT (bottom left graph) and control time 2 (bottom right graph) conditions. For each graph, each point represents one participant. Plotted points above the dashed line represent participants with higher AUCs in the condition labelled on the y-axis. Plotted points below the dashed line represent participants with higher AUCs in the condition labelled on the x-axis (i.e., control time 1). Note that the top left graph is the same as the top left graph in Figure 5.2; this graph is shown again for comparative purposes within the current figure.

Figure 5.3 also depicts the individual data, however the comparisons shown address the two major research questions of the current experiment (Do EPT and SFT reduce discounting rate? Does EFT have a lasting effect?). As shown in the top right graph, approximately half (52%) of the participants had higher AUCs in the EPT condition compared to the control condition, and 8% of participants had similar AUCs in each condition. As shown in the bottom left graph, fewer than half (48%) of the participants had higher AUCs in the SFT condition compared to the control condition, and 7% of participants had similar AUCs in each condition.

The bottom right graph of Figure 5.3 shows the AUCs for the control time 2 condition relative to the AUCs for the control time 1 condition. Most participants (63%) had higher AUCs in the control time 2 condition, and 5% of participants had similar AUCs in each condition. However, this difference was not statistically significant at the Bonferroni-corrected alpha level of 0.005 (see Table 5.2), and medians are very similar (0.55 vs. 0.57; see Figure 5.1).

The finding that EFT reduced delay discounting rate compared to the control time 1, EPT and SFT conditions replicates the general EFT effect, consistent with previous research (e.g., Benoit et al., 2011; Chiou & Wu, 2017; Dassen et al., 2016; O'Donnell et al., 2017; Peters & Büchel, 2010; Stein et al., 2017). Contrary to our predictions, the non-significant difference between the EPT and control conditions suggests that future thinking is necessary to reduce delay discounting rate. We predicted that EPT would have a similar effect to EFT on delay discounting rate, based on the temporal attention hypothesis, the CES hypothesis and findings by Lempert et al. (2017). These results suggest that, although thinking about the past may still be important in constructing an imagined future, thinking about the past does not appear to be sufficient on its own to reduce discounting rate.

Our result is consistent with Daniel et al. (2016) who concluded that EFT and EPT effects on delay discounting are not interchangeable. Our result is also consistent with additional unpublished research within our lab, where Walsh (2018) attempted to conduct a closer replication of Lempert et al.'s (2017) methodology, but still found no significant difference between EPT and NT conditions. However, the most notable difference in methodology that remains between the experiments within our lab (i.e., the current experiment and Walsh, 2018) and Lempert et al. is that participants in Lempert et al.'s study had a three-day delay between writing about their past events and completing the delay discounting task. It is possible that this three-day delay gave participants more time to reflect on their past events before completing the discounting task, therefore strengthening the effectiveness of the EPT manipulation. Regardless, all results taken together suggest that EPT alone is an unreliable intervention for reducing delay discounting rate.

Furthermore, the non-significant difference between the SFT and control conditions suggests that episodic thinking may also be necessary to reduce delay discounting rate. This result is consistent with Chiou and Wu (2017) and Wu et al. (2017). However, this result is inconsistent with Palombo et al. (2016) and Kwan et al. (2015), whose results suggest that SFT can reduce discounting rate. One limitation of the current experiment is that the SFT condition had a limited future component. That is, we did not prompt participants to identify items that they could purchase at specific delays in the future (as in Palombo et al., 2016), rather, participants were instructed to think of an item that a person could buy in the future in general. This is a limitation now that we know that future thinking is necessary to reduce delay discounting rate, and could therefore account for the difference in results. Thus, it could be that the SFT condition had no effect because there was essentially no future thinking component, *or* because there was no episodic thinking component. However, previous research suggests that the EFT effect does not depend on temporal specificity. Peters and Büchel (2010) found comparable EFT effects when participants thought about the future more generally (i.e., when participants were instructed to list events that were all “within the next couple of months” but not tied to a particular date) and when participants thought about the future with specific time points as in typical EFT studies. Additionally, both Chiou and Wu (2017) and Wu et al. (2017) described the future component of their SFT and EFT conditions in the same way (using one delay of “1 year”), yet only EFT reduced delay discounting in those studies. Thus, it is unlikely that the generality of the future component of the SFT condition accounts for the current results.

An alternative explanation for the different findings between Palombo et al. (2016) and the current study is that Palombo et al.’s SFT effect may be explained by demand characteristics. For example, Palombo et al. asked participants to imagine what they could buy with $42 in 6 months and then asked them to choose between $30 now and $42 in 6 months. Our SFT manipulation (as well as Chiou & Wu’s 2017 and Wu et al.’s 2017 SFT manipulations), did not completely overlap with content in the money task, therefore demand characteristics were less likely to contribute to the results. Indeed, few participants (9%) in the current study indicated that the researcher expected participants to be more likely to select the larger, delayed option in the SFT condition (see other demand characteristics analyses below). The finding that the SFT condition in the current study did not reduce discounting compared to the NT control also shows that the financial relevance component of EFT is not sufficient on its own to reduce discounting.

Finally, the non-significant difference between the control time 1 and control time 2 conditions suggests that there is no long-lasting effect of EFT on discounting rate. However, one difference between the control time 1 and control time 2 conditions is that the control time 2 condition was always experienced at the end of the session. If participants became more impulsive during the last condition, then this would cancel out any lasting EFT effect, and could therefore explain the non-significant difference between control time 1 and control time 2. In order to investigate whether there was an effect of time on the AUCs, we compared AUCs among conditions experienced first, second, third, fourth and fifth for each participant. Figure 5.4 shows the group median AUC for each condition.

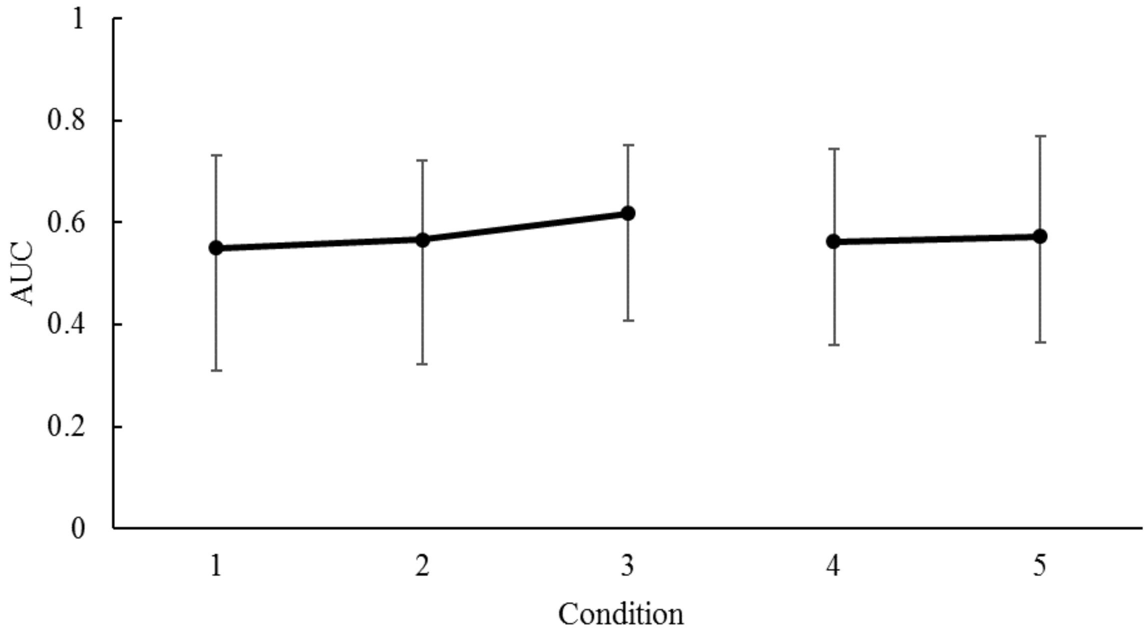

Figure 5.4. Group median AUC for conditions experienced first, second, third, fourth and fifth. The first and fifth conditions were always control conditions. During the second, third, and fourth conditions each participant completed EFT, EPT, and SFT in a random order. Participants completed the first, second and third conditions during the first session and the fourth and fifth conditions during the second session. Error bars represent the interquartile range.

A Friedman test showed a significant main effect of order on AUCs (X (4) = 10.62, p = 0.031). However, follow-up Wilcoxon signed-rank tests showed no significant differences in AUCs between conditions 1 and 2 (Z = -0.92, p = 0.358), conditions 2 and 3 (Z = -1.34, p = 0.181), conditions 3 and 4 (Z = -1.34, p = 0.181), or conditions 4 and 5 (Z = -0.533, p = 0.594). The first condition was always control time 1 and the fifth condition was always control time 2. The difference between these two conditions was also non-significant at the Bonferroni-corrected alpha level of 0.01 required for this analysis (see Table 5.2). Thus, the current results do not provide any evidence of a lasting effect of EFT on delay discounting rate, and order cannot account for this null result. Rather, it appears that participants need to be continuously cued for EFT to reduce delay discounting rate.

Another finding that task order could potentially account for is the puzzling non-significant difference between the EFT and control time 2 conditions. It is possible that the EFT manipulation had a carry-over effect into the control time 2 condition for participants who experienced the EFT condition in the second session (i.e., right before control time 2). Although this explanation seems unlikely given the results of Experiment 1 (where there was no evidence of a carry-over effect), it cannot be entirely ruled out. Thus, we investigated this possibility by comparing AUC results between participants who experienced the EFT condition in the first session and participants who experienced the EFT condition in the second session. If the EFT manipulation had a carry-over effect, then we would expect a significant difference between EFT and control time 2 for participants who experienced the EFT condition in session 1, but not for participants who experienced EFT in session 2.

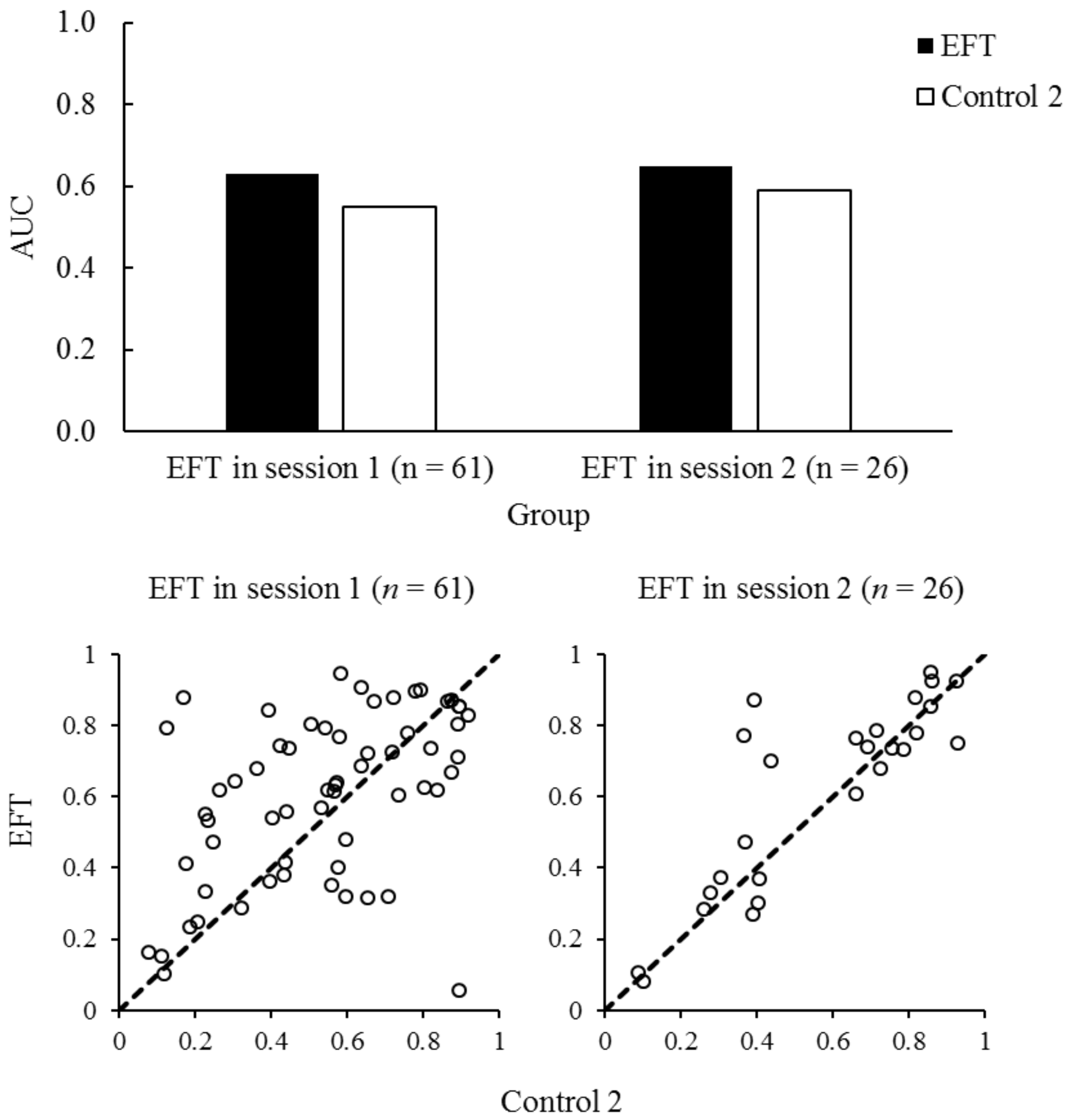

Figure 5.5. The top bar graph represents the group mean AUCs (trimmed by 20%) for the EFT (black bars) and control time 2 (white bars) conditions for participants who experienced EFT in session 1 and session 2. The bottom modified Brinley plots show the individual data for participants who experienced EFT in session 1 (left graph) and for participants who experienced EFT in session 2 (right graph). Plotted points above the dashed line represent participants with higher AUCs in the EFT condition, whereas plotted points below the dashed line represent participants with higher AUCs in the control 2 condition.

Figure 5.5 shows the AUCs for the EFT and control time 2 conditions for participants who experienced EFT in session 1 and for participants who experienced EFT in session 2. The top graphs represent the group means (trimmed by 20%) and the bottom graphs represent the individual data. A two-way mixed robust ANOVA (computed using the R function bwtrimbt, as described in Chapter 4) revealed a significant main effect of EFT ($p = 0.023$), no significant main effect of EFT session number ($p = 0.694$), and no significant interaction between EFT and EFT session number ($p = 0.854$). Thus, the session number in which participants experienced the EFT condition cannot account for the non-significant difference between EFT and control time 2. Although the main effect of EFT was statistically significant at an alpha level of 0.05, note that this was not statistically significant at the Bonferroni-corrected alpha level of 0.005 required for the overall results, and is therefore consistent with the Friedman test results (see Table 5.2)[14]. Nevertheless, the non-significant difference between EFT and control time 2 does not cast doubt on the replicability of the EFT effect, as the EFT effect was consistently found for three out of the four comparisons. That is, EFT reduced delay discounting rate compared to control time 1, EPT and SFT conditions, consistent with previous research (e.g., Benoit et al., 2011; Chiou & Wu, 2017; Dassen et al., 2016; O'Donnell et al., 2017; Peters & Büchel, 2010; Stein et al., 2017).

The size of the EFT effect as a function of delay

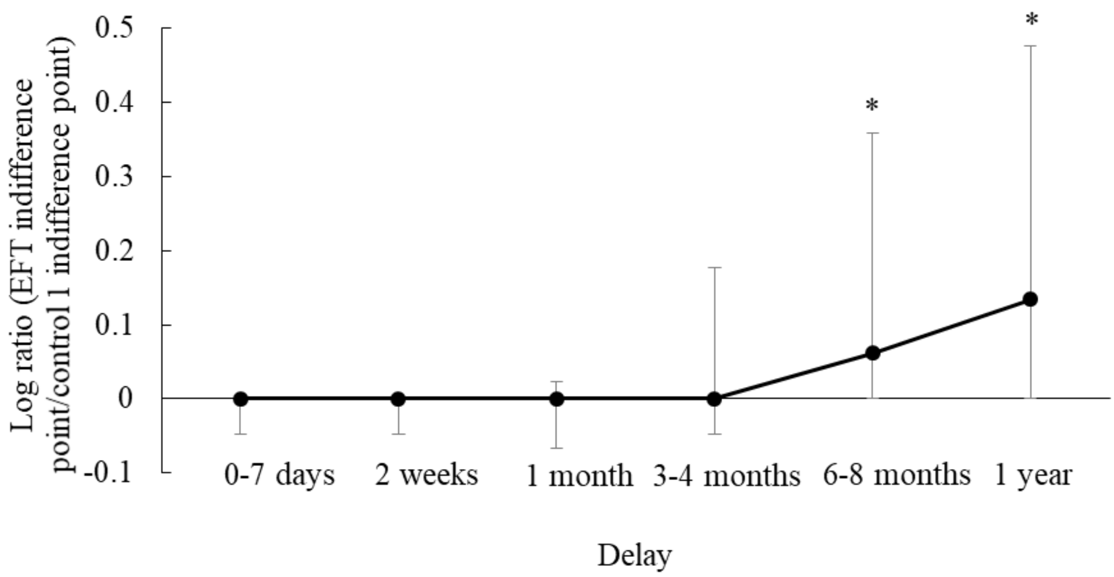

Figure 5.6. The group median size of the EFT effect (log [EFT indifference point/control 1 indifference point]) as a function of delay. Points above the x-axis indicate that participants were more self-controlled during the EFT condition than during the first control condition. * indicates that the size of the EFT effect was significantly different from zero. The error bars represent the interquartile range.

A Friedman test showed a significant main effect of delay on the size of the EFT effect (log [EFT indifference point/control 1 indifference point]), $X = 74.50$, $p < 0.001$. As

[14] A robust one-way repeated measures ANOVA (computed using the R function rmanovab which uses trimmed means and a bootstrapping procedure; see Chapter 4 for further explanation) also produced results consistent with the original Friedman test.

shown in Figure 5.6, EFT only reduced discounting at the 6-8-months and 1-year delays. One sample Wilcoxon signed rank tests confirmed that the size of the EFT effect was significantly different from zero at the 6-8-months ($p = 0.002$) and 1-year ($p < 0.001$) delays. There was no significant difference in the size of the EFT effect between the 6-8-months and 1-year delays ($Z = -1.22$, $p = 0.221$).

Manipulation check questions

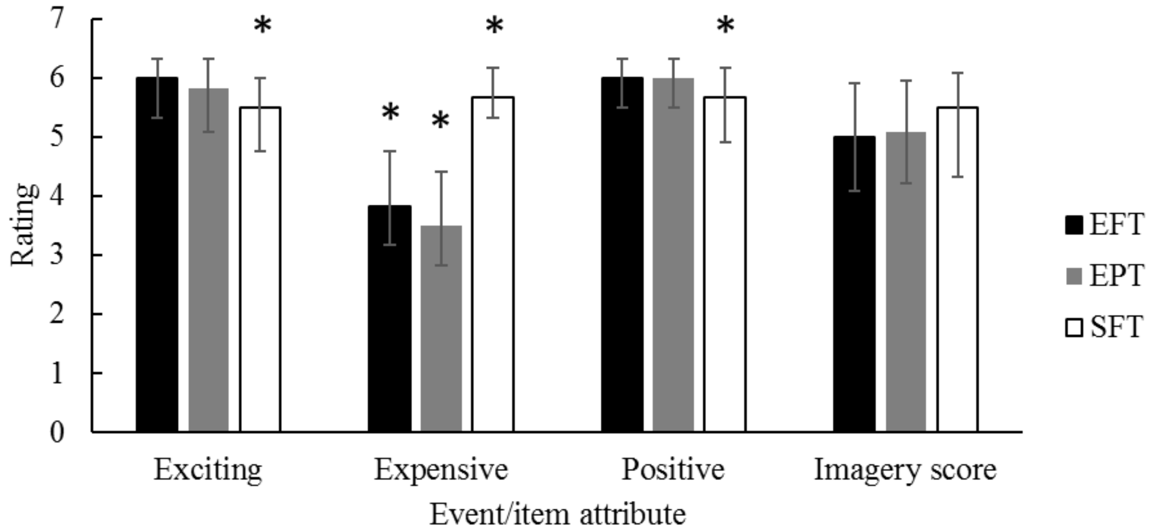

Figure 5.7. Group median ratings for how exciting, expensive and positive events/items were in the EFT (black bars), EPT (grey bars) and SFT (white bars) conditions. The group median imagery score (averaged frequency and vividness of imagined events/items during the money task) is also shown for each condition. * indicates that the rating for that condition is significantly different from the ratings for all other conditions for that particular attribute. Error bars represent the interquartile range.

Figure 5.7 shows how exciting, expensive and positive events/items were for each condition, as well as the imagery scores. Overall, ratings were high across each event/item attribute for each condition, except for expensiveness ratings for the EFT and EPT conditions, which were below the midpoint of the scale. The items from the SFT condition were rated as significantly less exciting, less positive, and more expensive compared to the events from EFT and EPT conditions (all ps < 0.002). The events in the EFT condition were rated as significantly more expensive than the events in the EPT condition ($p < 0.001$), however there were no significant differences between the EFT and EPT conditions for any other attribute (all ps > 0.05). There was no significant main effect of condition for the imagery scores (X (2) = 5.70, $p = 0.060$). The initial Friedman tests for the other attributes are in Appendix J.

These results show that differences in degree of visual imagery cannot account for the significant differences between EFT AUCs and EPT/SFT AUCs. These results also show that differences in how exciting, expensive or positive events/items were in each condition cannot entirely account for the AUC results, given that ratings were either in the opposite direction of the AUC results (e.g., SFT items were rated as the most expensive), or could not account for all significant differences in AUCs (e.g., while the SFT items were rated as significantly less positive, there was no significant difference in positive ratings between the EFT and EPT conditions).

Participants rated the event at the indifference point with the largest EFT effect (log [EFT indifference point/control 1 indifference point]) as significantly more expensive (*Mdn* = 5) than the event at the indifference point with the smallest EFT effect, (*Mdn* = 4) $Z = -2.56$, $p = 0.010$. These two indifference points did not significantly differ on any other event attribute (imagery scores: $Z = -0.96$, $p = 0.335$; all other medians were identical). Overall, these findings are consistent with previous research which showed that although financial relevance cannot completely account for the EFT effect, it may amplify the effect (O'Donnell et al., 2017).

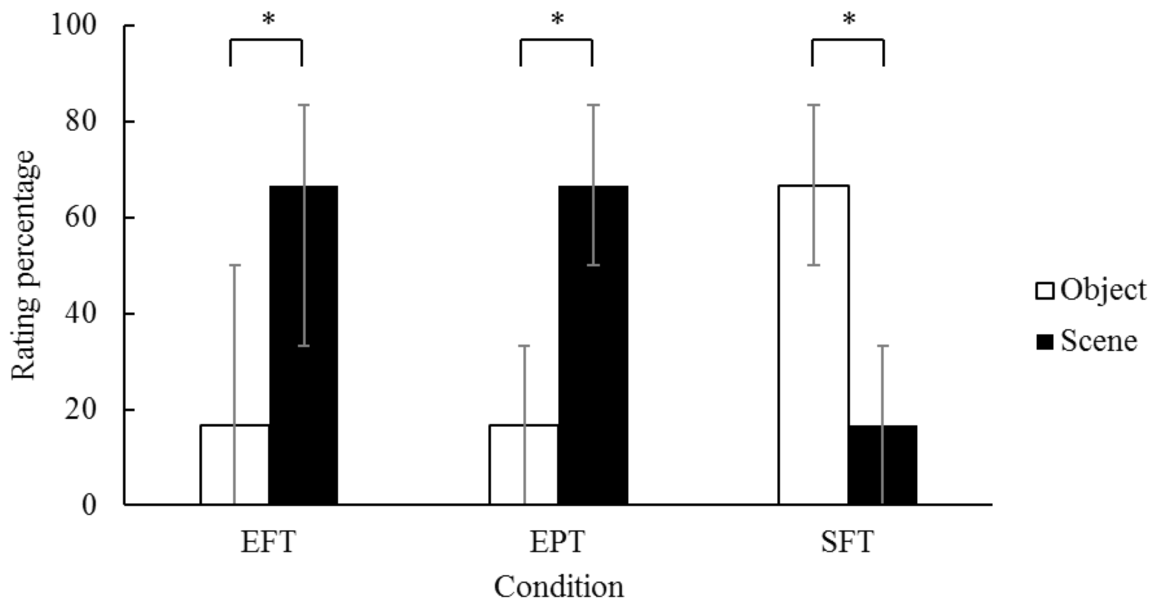

Figure 5.8. Group median percent of instances where participants chose each rating type for the EFT, EPT and SFT conditions. Black bars represent 'scene' ratings and white bars represent 'object' ratings. * indicates a significant difference between the two bracketed conditions. Group median percent of instances where participants chose the 'nothing' rating

was zero in each condition, therefore this rating type is not shown. Error bars represent the interquartile range.

Figure 5.8 shows the group median percent of instances where participants stated they were thinking about scenes or objects for each condition. In both the EFT and EPT conditions, participants were thinking about a scene unfolding significantly more than they were thinking about objects (EFT: $Z = -4.81$, $p < 0.001$; EPT: $Z = -6.32$, $p < 0.001$). Conversely, in the SFT condition, participants were thinking about objects significantly more than they were thinking about a scene unfolding ($Z = -6.13$, $p < 0.001$). Between-condition comparisons showed that participants thought about objects significantly more (and a scene unfolding significantly less) in the SFT condition compared to the EFT and EPT conditions (all ps < 0.001). There was no significant difference in scene or object ratings between the EFT and EPT conditions (both scene medians = 66.67; both object medians = 16.67). These results show that the episodic/semantic manipulations were successful, and are comparable to the percentage of scene/object ratings found in Palombo et al. (2016). Although participants were thinking about very different things during the EPT and SFT conditions (EPT = scenes; SFT = objects), they both had the same, non-significant effect on discounting rate. This further shows that episodic thinking is not sufficient on its own to reduce discounting rate; future thinking is a necessary component.

In order to assess whether the EFT effect was stronger for participants who thought about a scene unfolding more often, we calculated a Spearman's rank correlation between the size of the EFT effect (log ratios of EFT and control time 1 AUCs for each participant) and the percentage of times participants said that they thought about a scene unfolding during the EFT condition. There was no significant correlation between log ratios and percent scenes, r_s (89) = 0.07, $p = 0.534$. Thus, participants who thought about scenes more often during EFT were not more likely to have a larger EFT effect. Figure 5.9 shows the log ratios plotted against the percent scenes.

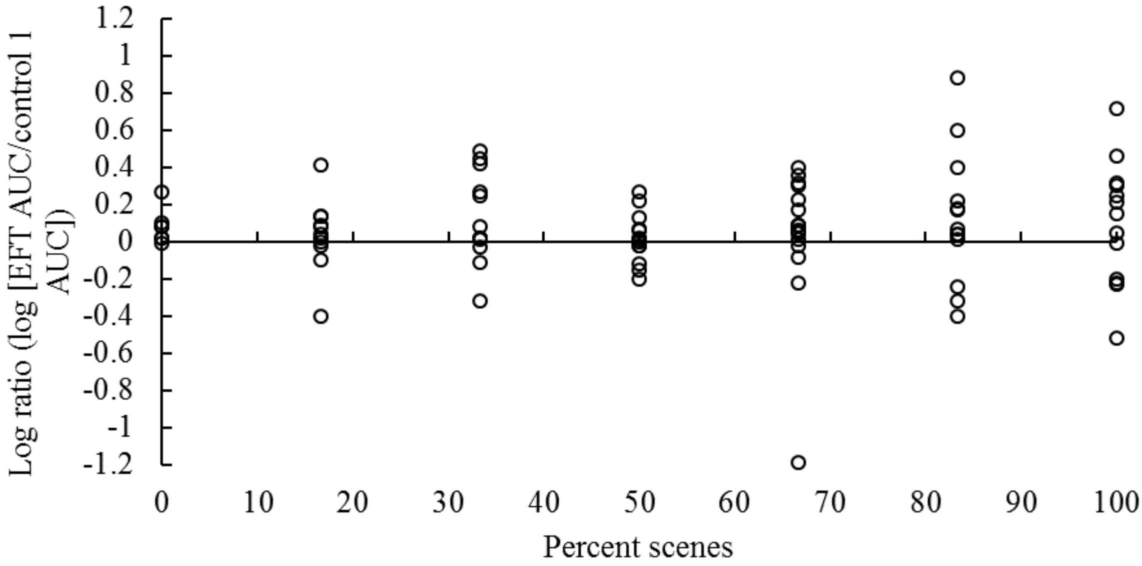

Figure 5.9. The size of the EFT effect (log [EFT AUC/control 1 AUC]) plotted against percent scenes. Points above the x-axis are participants who were more self-controlled during the EFT condition than during the first control task.

We also calculated the log ratio of EPT and control time 1 AUCs and the log ratio of SFT and control time 1 AUCs for each participant. Again, there was no significant correlation between log ratios and percent scenes for the EPT, r_s (89) = -0.14, p = 0.207, or for the SFT, r_s (89) = 0.05, p = 0.679, conditions.

Comparison of event ratings between EFT indifference points with the smallest and largest deviations from the hyperbolic model

We also found that participants rated the event at the indifference point that positively deviated the most from the hyperbolic model as significantly more exciting (*Mdn* = 7) than the event at the indifference point that deviated the least from the hyperbolic model, (*Mdn* = 6), Z = -2.43, p = 0.015. These two indifference points did not significantly differ on any other event attribute according to a Bonferroni-corrected alpha level of 0.016 (expensiveness: largest deviation *Mdn* = 5, smallest deviation *Mdn* = 4, Z = -2.33, p = 0.020; imagery score: largest deviation *Mdn* = 5, smallest deviation *Mdn* = 5.5, Z = -0.06, p = 0.949; all other medians were identical). This result supports the conclusion that the high levels of unsystematic discounting in the EFT condition are due to the different events being unmatched in excitingness.

Reaction times

Thus far, we have concluded that the EFT effect is caused by the *type* of thinking required during the EFT condition (i.e., future thinking). However, it is possible that the EFT effect is instead caused by the *degree* of thinking required during the EFT condition. Indeed, people have reported that future events are more difficult to imagine than past events (Özbek, Bohn & Berntsen, 2017), indicating that a more taxing thought process may be required during EFT. A higher cognitive load in the EFT condition could also explain the high levels of unsystematic discounting found within this condition (Franco-Watkins et al., 2010). In order to assess whether degree of thinking was greater in the EFT condition, we compared participants' reaction times (i.e., the amount of time that they spent completing the money task) among conditions (see Figure 5.10). If people took longer to complete the money task during the EFT condition, then this would suggest that the EFT effect is caused by the EFT condition requiring a more taxing thought process (i.e., a higher *degree* of thinking), rather than by the process of thinking about the future (i.e., the *type* of thinking involved).

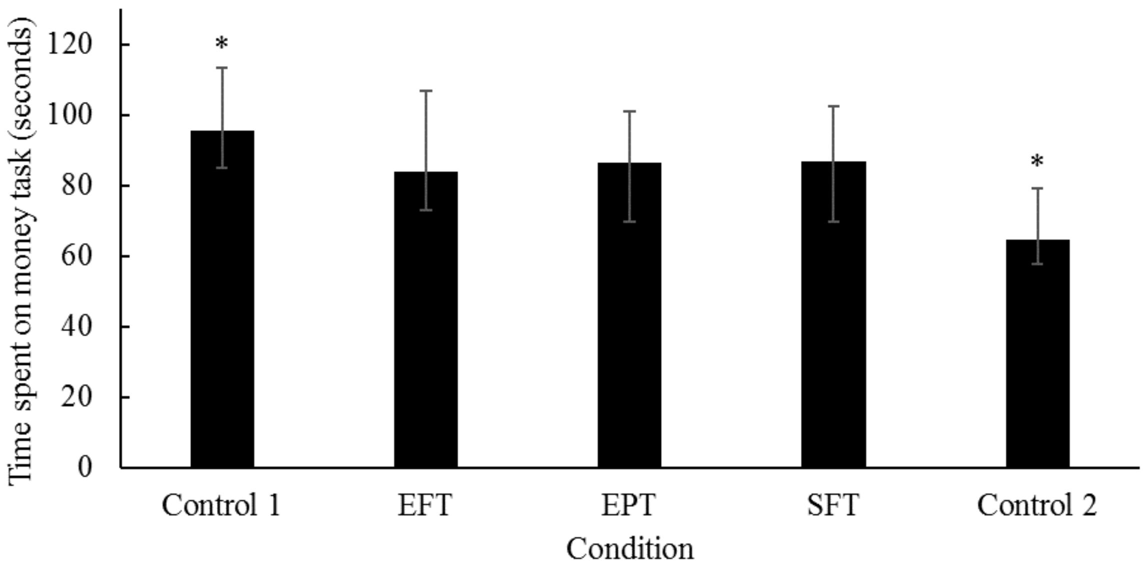

Figure 5.10. Group median amount of time that participants spent completing the money task during each condition. * indicates that the amount of time spent during that condition is significantly different from all other conditions. Error bars represent the interquartile range.

A Friedman test showed a significant main effect of condition on time spent on the money task (X (4) = 126.27, $p < 0.001$). Follow-up Wilcoxon signed-rank tests revealed that participants took significantly longer on the money task during the control time 1 condition compared to all other conditions (all *p*s < 0.001). Participants also spent significantly less time on the money task during the control time 2 condition compared to all other conditions (all *p*s < 0.001). All other comparisons were non-significant (all *p*s > 0.05). These results suggest that EFT did not produce a more taxing thought process than EPT or SFT. Thus, the EFT effect is produced by the *type* of thinking required (i.e., future thinking), rather than the *degree* of thinking required, during the EFT condition. These results also show that participants may have become faster at completing the money task over time, given that control time 1 was always experienced first and control time 2 was always experienced second. Figure 5.11 shows the effect of condition order on time spent completing the money task.

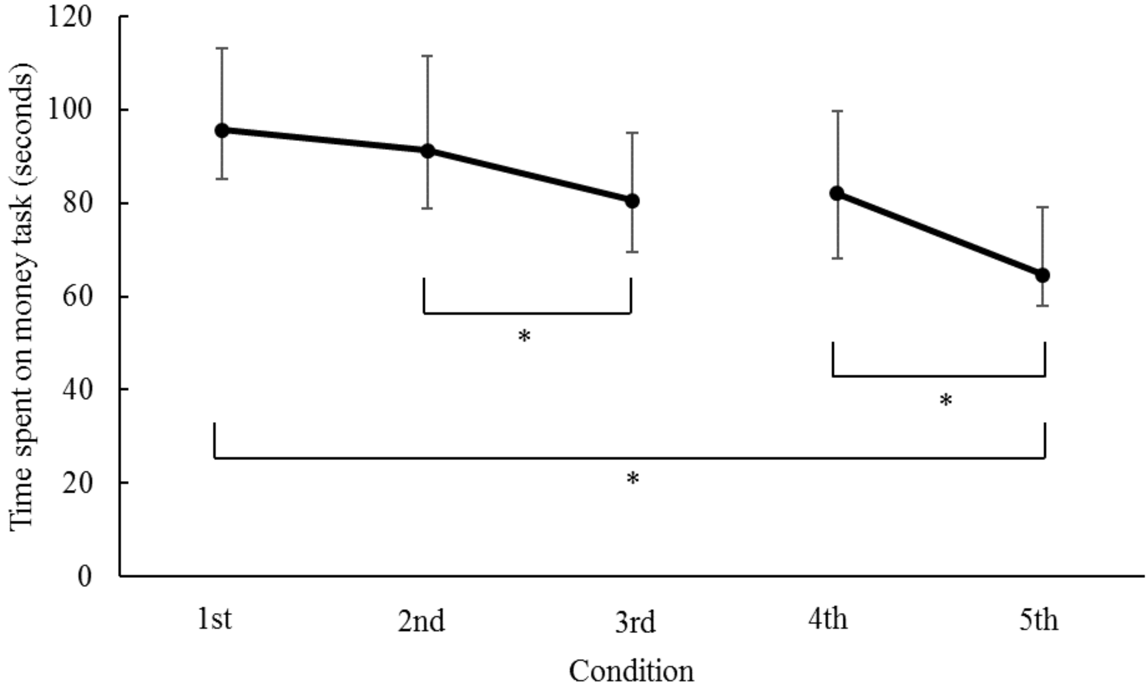

Figure 5.11. Group median amount of time that participants spent completing the money task during conditions experienced first, second, third, fourth and fifth. The first and fifth conditions were always control conditions. During the second, third, and fourth conditions each participant completed EFT, EPT, and SFT in a random order. Participants completed the first, second and third conditions during the first session and the fourth and fifth conditions during the second session.* indicates a significant difference between the two bracketed conditions. Error bars represent the interquartile range.

A Friedman test showed a significant main effect of order on time spent completing the money task (X (4) = 151.43, $p < 0.001$). Follow-up Wilcoxon signed-rank tests revealed no significant differences between the first and second conditions ($Z = -0.65$, $p = 0.516$) or between the third and fourth conditions ($Z = -0.63$, $p = 0.528$). All other differences were significant (all *p*s < 0.001). These results confirm that participants responded faster during the money task over time, consistent with practice effects (Greenwald, 1976). However, it is also possible that participants took the task less seriously as time progressed.

Demand characteristics

As shown in Table 5.3, only a small number of participants (15%) correctly predicted the results of the current experiment, indicating that demand characteristics cannot account for the results. In fact, a larger percentage of participants (39%) incorrectly guessed the researcher's predictions and results. These participants guessed that either: 1) thinking about events/items would make people want the money sooner rather than later, 2) thinking about events/items will have no effect on decisions about money, or 3) that the money task served as a distractor task to see how well people could still think about their events or items. Twenty-two percent of the participants correctly stated the research question without specifying the direction of the hypothesis (e.g., that events will affect decisions about money). We are therefore unable to conclude whether 22% of the participants knew the direction of the researcher's hypothesis.

If we are conservative and consider the number of participants who made statements that did predict the group results, as well as the number of participants who *might* have predicted the group results, then this would suggest that only 37% of participants' predictions did (or might) match the overall results. This percentage is comparable to the percentage of participants who incorrectly guessed the researcher's predictions. Furthermore, only 12% of participants did predict (or could have predicted) the group results *and* had data showing that EFT reduced their discounting compared to control 1 (EFT AUC > control 1 AUC). Overall, these results show that demand characteristics are an unlikely explanation for the results of the current experiment.

Table 5.3
Participants' answers to the questions assessing demand characteristics coded into categories

Type of coded answer	Percentage of participants
Guessed that the researcher was interested in how people make decisions about money, with no mention of the cues	10% (8/82)

Incorrectly guessed the researcher's predictions and results	39% (32/82)
Stated that they did not know the researcher's predictions	1% (1/82)
Correctly guessed the researcher's predictions, but inconsistent with results (e.g., EFT *and* EPT will reduce discounting)	13% (11/82)
Guessed researcher's prediction that is consistent with results (i.e., only EFT will reduce discounting)	15% (12/82)
Correctly identified research question, but did not clarify direction of prediction (e.g., investigated whether EFT will affect decisions about money)	22% (18/82)

Conclusion

The current Experiment 3A investigated: 1) whether the episodic and future thinking components of EFT are necessary to reduce discounting rate, and whether one of these components is more important than the other, and 2) whether EFT has a lasting effect on discounting rate over time. The results of Experiment 3A replicated the general EFT effect, and also showed that brief EFT interventions are best suited for one-off decisions, given that EFT had no significant lasting effect on discounting rate. Thirdly, EPT and SFT had no significant effect on discounting rate. These results suggest that both future thinking and episodic thinking are necessary components of the EFT effect. We also showed that these results cannot be explained by the future events being more positive, expensive, exciting or vivid compared to the past events and semantic items. These results also cannot be explained by future thinking requiring a more taxing thought process compared to the past thinking and semantic thinking manipulations. Furthermore, the results cannot be explained by demand characteristics. Although our limited future thinking component in the SFT condition is an unlikely alternative explanation for the non-significant SFT effect, it cannot be completely ruled out. We were also unable to make a clear conclusion from the meta-analysis regarding the relative importance of the episodic thinking component, given that most studies only employed NT controls. Future research should attempt to design a SFT condition with a clear future component, while also accounting for potential demand characteristics.

Experiment 3B

Experiment 3B investigated whether another component of EFT is necessary to reduce delay discounting: personal relevance. Findings that highly personally-/emotionally-relevant events are correlated with larger EFT effects (Benoit et al., 2011) suggest that personal relevance might be an important component to the EFT effect. However, no research to date has manipulated the personal relevance of future events in order to clarify the importance of this component. In the current experiment, participants were randomly allocated to a personally relevant EFT condition where participants were instructed to think about their own personal future events (as in typical EFT studies), or a non-personally relevant EFT condition, where participants were instructed to imagine someone else's future events. Both groups also completed a control 'no thinking' condition.

The EFT conditions used in the current experiment were patterned after that of Yi, Pickover, Struppy-Sullivan, Baker, and Landes (2016) and Yi et al. (2017). Yi and colleagues have used EFT conditions where participants were required to imagine events from a list provided by the researcher, rather than generate their own events. For example, a participant might have been asked to imagine that they are engaging in a leisurely activity in one year. Participants were then asked to answer a set of questions regarding each event to encourage episodic thinking about the event, such as 'what will this activity be', 'where will it take place', 'who else will be there', etc. Yi et al. (2017) found that this EFT condition reduced delay discounting (compared to a control episodic present thinking condition), as have other researchers who have taken a similar approach and given participants a pre-determined list of events (e.g., Bulley et al., 2019; Palombo et al., 2015).

Yi et al. (2016) also manipulated episodic thinking about another person using this particular list of events, where participants were asked to imagine someone they know partaking in those listed events. For example, participants were asked to imagine that the person is engaging in a leisurely activity today, and to describe where it will take place and what time it will take place, etc. However, Yi et al. assessed the effect of episodic (non-future) thinking about others on social discounting, rather than on delay discounting. Social discounting refers to the phenomenon that the more socially close people feel towards another person, the more they are willing to forgo a reward for themselves to benefit that person. For example, people will forgo more money for themselves when it benefits a close friend, compared to if it benefits a stranger (Jones & Rachlin, 2006). Across a series of experiments, Yi et al. found that episodic thinking about others, as well as episodic future thinking about the self, reduced social discounting. However, whether episodic future thinking about others reduces delay discounting remains to be determined, hence the current study. Findings that delay discounting is theoretically and empirically associated with social discounting suggest that thinking about others may also reduce delay discounting. Rachlin and colleagues (Rachlin & Jones, 2009; Rachlin & Locey, 2011) have proposed that people might perceive their future self as a different person from their present self. Indeed, people make future decisions for themselves as if they were making them for another person (Pronin, Olivia, & Kennedy, 2008), and the more similar people view their current selves to their future selves, the less delay discounting they show (Bartels & Rips, 2010). Degree of social discounting is also positively correlated with degree of delay discounting across individuals (Rachlin & Jones, 2008; Jones & Rachlin, 2009; Yi, Carter & Landes, 2012).

We manipulated EFT in the current study using Yi and colleagues' (2016, 2017) list of events to ensure that the future events would be standardised across the personally relevant and non-personally relevant EFT conditions. Given that Yi et al. (2016, 2017) provided four different types of events in their studies, we used four delays in the current experiment, rather than six delays as in previous experiments of the current thesis. Because we had a smaller number of delays, we ensured that we chose a range that are known to produce EFT effects. Experiment 1 of the current thesis showed no significant EFT effect at delays shorter than one month. Snider et al. (2016) also found no significant EFT effect at delays shorter than one month, whereas Stein et al. (2016) found no significant EFT effect at delays shorter than three months. We therefore chose to use a longer range of delays consisting of three months, six months, one year and two years.

If participants in both the personally relevant and non-personally relevant groups have an EFT effect in the current study then this would suggest that thinking about a future episode is sufficient to reduce delay discounting, regardless of personal relevance. If the EFT effect is limited to (or larger in) the personally relevant group, then this would suggest that personal relevance is a necessary (or contributing) component of the EFT effect. Alternatively, the latter results could also arise because demand characteristics are more likely to contribute to an EFT effect in the personally relevant group. That is, the researcher's prediction (that future cues will cause people to choose the larger, delayed reward) might be more obvious to participants in the personally relevant condition because being instructed to think about their own future, rather than someone else's, better aligns with the money task where they are making decisions for themselves. We therefore assessed whether demand characteristics can explain any of our results by asking participants what they thought the research question and hypotheses were for the current experiment, as in Experiment 3A.

As previously noted in the general introduction, EFT might reduce discounting by making the larger, delayed reward more valuable or by making the delay to the larger, delayed reward seem shorter. We therefore included a multiple-choice question to determine whether participants perceived EFT to reduce their discounting through one or both of these mechanisms.

Method

Design

We used a mixed design where EFT (EFT vs. control) was manipulated within subjects and personal relevance (personally relevant EFT vs. non-personally relevant EFT) was manipulated between subjects.

Participants

Participants were 127 third-year psychology students at Victoria University of Wellington. No other demographic information was collected. Participants completed the following tasks during their tutorial as a compulsory course requirement. However, choosing to include their data for research purposes was optional, and informed consent was obtained prior to participation. Participants were randomly assigned to the personally relevant condition (N = 57) or the non-personally relevant condition (N = 70).

Materials

Participants completed the following tasks in the laboratory on computers. The tasks were constructed using Microsoft Visual Basic® 2017 Software.

EFT manipulation for the personally relevant group. Participants were asked to imagine four specific future events involving themselves (having lunch, visiting a website, engaging in a leisurely activity, and having a conversation; as in Yi et al., 2016). Each of the four types of events was randomly paired with one of four delays (three months, six months, one year and two years). For example, a participant could be asked to imagine that they are having lunch in three months, having a conversation in six months, visiting a website in one year, and engaging in a leisurely activity in two years. Participants were required to type answers to six questions regarding each specific event, which were designed to evoke episodic future thinking (see Appendix K for all questions for each event type). Prior to answering these questions, participants were told that they should consider what their life could be like at this particular time in the future, and that the details regarding the event should be plausible (based on what their life could be like at this time), emotionally positive, personally relevant and exciting. After answering the six questions, participants typed a cue word or phrase for the event.

EFT manipulation for the non-personally relevant group. The EFT manipulation for the non-personally relevant group was identical to the personally relevant group, except that participants were asked to think about someone else’s future events, rather than their own. Participants were asked to think about a person they know well, such as a close friend or relative, and to type that person’s name into the text box provided so that this name could be incorporated into the six questions for each event (see Appendix K). For example, if a participant entered the name ‘Alex,’ then they would subsequently be asked to imagine Alex engaging in a leisurely activity in X delay. They would also be instructed to consider what Alex’s life could be like at this particular time in the future, and that the details regarding the event should be plausible, emotionally positive, personally relevant and exciting for Alex. Participants were also told that they themselves would not be present during the events or involved in any way.

Control ‘no thinking’ delay discounting task. The control delay discounting task was the same as the control delay discounting task from Experiment 3A (where participants were prompted to ‘choose’), except that the delays were three months, six months, one year and two years.

EFT delay discounting task. The EFT delay discounting task was identical to the control delay discounting task except that: 1) the cue word or phrase for the event corresponding to each delay was written in red centred above the two choices, rather than the word ‘choose’, and 2) the instructions prompted participants to imagine the future events corresponding to the cues. The instructions for the personally relevant group were identical to those in the EFT condition from Experiment 3A. The instructions for the non-personally relevant group were identical to the personally relevant group except that participants were instructed to imagine that the event is really happening for the person they previously wrote about.

Manipulation check questions. As in Experiment 3A, participants rated how often each cue made them think about the associated imagined event during the money task, how vivid, positive, and exciting the imagined event was, how expensive the imagined event would be, and whether each cue made them think about scenes, objects or nothing. In addition to rating how positive and exciting imagining the future events were for themselves ('self' ratings), participants in the non-personally relevant group also rated how positive and exciting the events would be for the specified person ('other' ratings). Both groups also rated how personally relevant each imagined event was (1 = not personally relevant to me at all, 7 = highly personally relevant to me).

Finally, participants were asked the following open-ended questions: 1) *What did you think about when you answered the questions about money when you were prompted to imagine events?*, and 2) *What did you think about when you answered the questions about money when you were simply prompted to 'choose'?*

Questions to assess demand characteristics. The questions to assess demand characteristics were identical to those in Experiment 3A. We also asked participants if they had discussed this experiment with a classmate who completed it earlier in the week, in order to assess if they already knew the purpose of the experiment prior to participation. All participants stated that they had not discussed the experiment.

Questions to assess EFT mechanisms. Participants were asked the following open-ended question: *Sometimes, imagining future events has an effect on the types of decisions that people make. Did thinking about future events affect your choices about money during this experiment? How?* Participants typed their answer, and then proceeded to the following multiple-choice question: *Did thinking about future events affect your choices about money during this experiment? Please select one of the following: a) Yes - thinking about future events made me more likely to choose the larger, delayed amount of money, b) Yes - thinking about future events made me more likely to choose the smaller, immediate amount of money, c) No - thinking about future events had no effect on my choices about money.* If participants selected 'a)', then they proceeded to an additional multiple-choice question: *Why did thinking about future events make you more likely to choose the larger, delayed amount of money?*

Please select one of the following: a) Thinking about future events made that future amount of money seem more valuable, b) Thinking about future events made the wait for the money seem shorter, c) Both 'a' and 'b', d) I don't know why thinking about future events made me more likely to choose the larger, delayed amount of money.

Procedure

All participants completed the tasks in the following order: EFT manipulation (either personally relevant or non-personally relevant), delay discounting tasks (with the order of the EFT and control tasks counterbalanced), manipulation check questions, questions to assess demand characteristics, questions to assess EFT mechanisms. Finally, participants were given a written and verbal debrief which explained the purpose of the experiment. The entire experiment took approximately 30 minutes to complete.

Unsystematic discounting and data exclusion

As in previous experiments, we excluded participants with unsystematic discounting in the control condition according to Johnson and Bickel's (2008) criteria (personally relevant group n = 1; non-personally relevant group n = 3). Thus, the final analysed sample was 56 in the personally relevant group and 67 in the non-personally relevant group. Table 5.4 shows the percentage of participants with unsystematic discounting (based on Johnson & Bickel's 2008 criteria) in each condition in each group.

Table 5.4
Percentage of participants with unsystematic discounting (based on Johnson & Bickel's 2008 criteria) in the EFT and control conditions for the personally relevant and non-personally relevant groups.

	Personally relevant group			Non-personally relevant group		
Condition	Criterion 1 only	Criterion 2 only	Both criteria	Criterion 1 only	Criterion 2 only	Both criteria
EFT	0%	1%	11%	4%	11%	7%
Control	0%	2%	0%	0%	4%	0%

One additional participant from the personally relevant group did not finish the entire experiment and therefore all qualitative analyses (i.e., answers to the "what did you think about" open-ended questions, questions to assess demand characteristics, questions to assess EFT mechanisms) are based on 55 (rather than 56) participants from this group. However, all 56 participants from the personally relevant group contributed to all other analyses. All participants in the non-personally relevant group (n = 67) completed all tasks and therefore contributed to all analyses.

Data analyses

Given that participants did not produce unreasonably high levels of unsystematic discounting (see Table 5.4), we intended to use k values (as well as AUC values) as a discounting rate measure. Thus, the hyperbolic model was fit to each participant's indifference points for each discounting task condition, as in previous experiments. However, the range of R^2 values indicated that the hyperbolic model did not describe the data well, even after unsystematic participants had been removed. Approximately one quarter (24%) of the participants in the personally relevant group had R^2 values lower than 0.5 in the EFT condition, and 17% of participants in the non-personally relevant group also had these low R^2 values in the EFT condition. 16% and 9% of the participants also had poor R^2s in the control conditions in the personally relevant and non-personally relevant groups, respectively. Additionally, examination of the residuals showed that the model underestimated value at the shorter delays and overestimated value at the longest delay in the control condition for the personally relevant group (see Appendix I). Therefore, only AUC values were used as a discounting rate measure.

Tests of normality (Shapiro Wilk and visual inspection of histograms) showed that the AUC values were normally distributed in the EFT and control conditions for both groups (all ps > 0.05). We therefore analysed the effect of EFT on AUC using a three-way mixed ANOVA (EFT x group x order). Significant effects were followed up with post-hoc tests using a Bonferroni-corrected alpha.

We calculated the percent of instances in which participants thought about a scene unfolding, objects, or nothing for each cue, as in Experiment 3A. Given that the group medians for the percent 'nothing' ratings were zero, only percent 'scenes' were compared between the two groups. These data, as well as all other ratings from the manipulation check questions, were not normally distributed (all ps > 0.05). We therefore compared ratings between the personally relevant and non-personally relevant groups using Mann Whitney U tests. We used a Bonferroni-corrected alpha level of 0.006 due to the high number of comparisons (8).

We also compared the event ratings between the indifference points with the largest EFT effect and the indifference points with the smallest EFT effect using Wilcoxon signed-rank tests, as in Experiment 3A. The size of the EFT effect at each indifference point was determined by calculating the log ratio of the EFT indifference point and the control indifference point (log [EFT indifference point/control indifference point]). We also compared the event ratings between EFT indifference points that positively deviated the most from the hyperbolic model and EFT indifference points that deviated the least from the hyperbolic model, as in Experiments 1 and 3A.

For the answers to the “what did you think about” questions, we coded whether participants stated that they thought about the events during the EFT and control discounting tasks. We also coded whether participants stated that they thought about the cost of the events. We calculated the percentage of participants that made these statements in the personally relevant and non-personally relevant groups, and compared these percentages between the two groups using Chi-Square tests. When expected frequencies were less than five, the *p* value from Fisher’s exact test, rather than from Pearson’s Chi-square, is reported (Field, 2009). An effect size (odds ratio; OR) was calculated for significant results using Equation 5.1 (Field, 2009):

$$OR = \frac{\left(\frac{N_1\, statement}{N_1\, no\, statement}\right)}{\left(\frac{N_2\, statement}{N_2\, no\, statement}\right)} \qquad \text{Equation 5.1}$$

Where N_1 *statement* is the number of participants in the personally relevant group who made the statement, N_1 *no statement* is the number of participants in the personally relevant who did not make the statement, N_2 *statement* is the number of participants in the non-personally relevant group who made the statement, and N_2 *no statement* is the number of participants in the non-personally relevant group who did not make the statement.

Participants’ answers to both of the demand characteristics questions were collated into one answer and coded into categories, as in Experiment 3A. We calculated the percentage of participants from the personally relevant and non-personally relevant groups with each type of coded answer, and made comparisons between the two groups using a Chi-square test. We also calculated the percentage of participants who predicted that the researcher expected EFT to reduce discounting *and* had data showing that EFT reduced their discounting. We defined EFT-reduced discounting at the individual level as an AUC in the EFT condition that was larger than an AUC in the control condition by more than zero, as in Experiment 3A. Although this criterion may have overestimated the percentage of individuals who had a meaningful change in discounting, it provided a strong test for demand characteristics.

In order to determine whether the size of the EFT effect was related to the mechanism through which participants reported they became more self-controlled, we calculated the percentage of participants who selected each answer in response to the EFT mechanisms multiple-choice question. Only participants who had accurately stated that EFT reduced their discounting (EFT AUC > control AUC) were included in this analysis. We also calculated the size of the EFT effect (log [EFT AUC/control AUC]) for each participant. These log ratios were not normally distributed (all Shapiro Wilkes *p*s < 0.05). We therefore used a Mann Whitney U test to determine if the size of the EFT effect was larger for participants who stated that EFT worked by both reducing perceived delay and increasing perceived value than for participants who stated that EFT worked by only increasing perceived value. We could not make comparisons with participants who stated that EFT only worked by reducing perceived delay because there were too few participants categorised into this group (see Results and Discussion).

Results and Discussion

Discounting rates

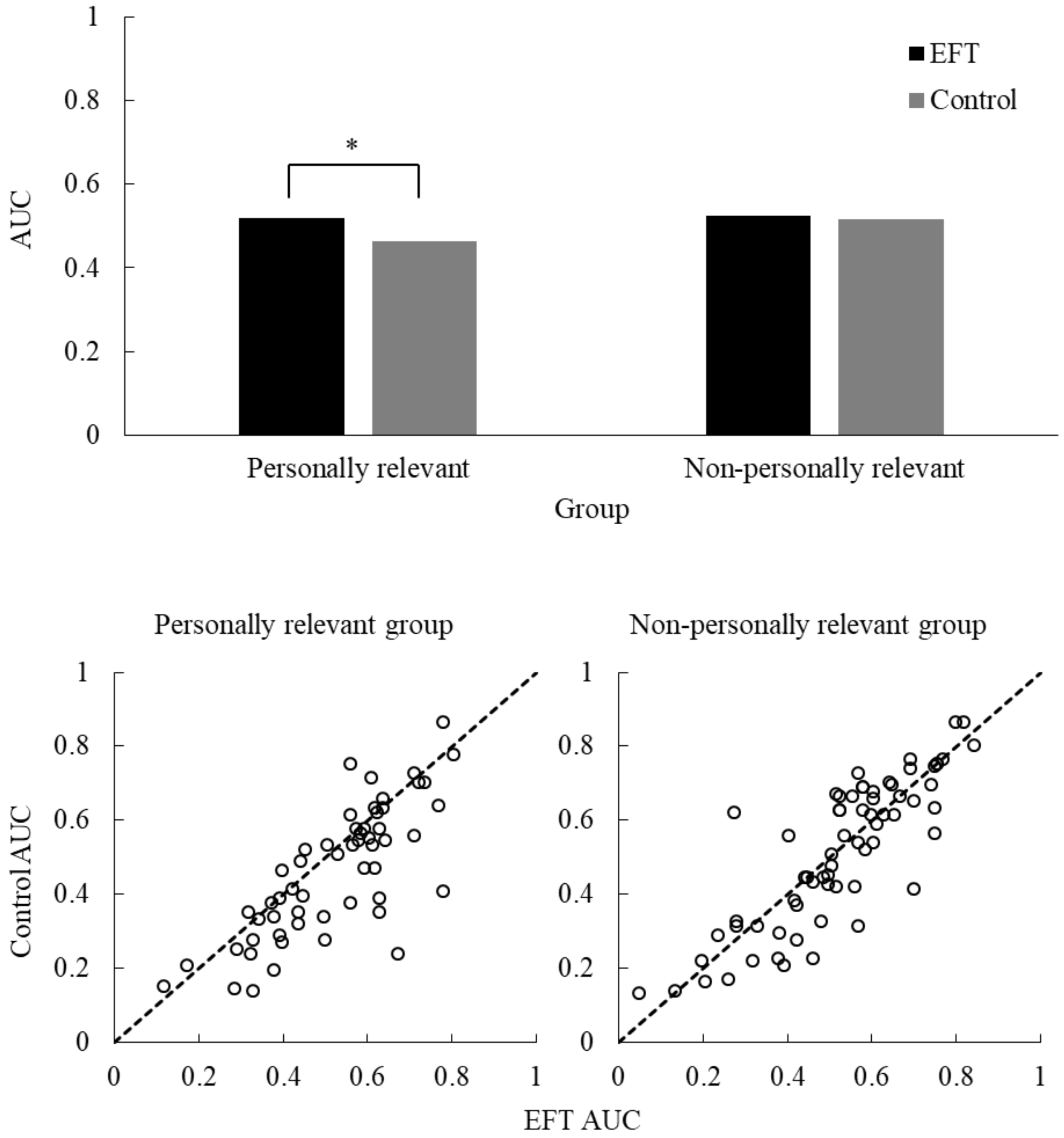

Figure 5.12. The top bar graphs show the group mean AUCs for EFT (black bars) and control (grey bars) conditions for the personally relevant and non-personally relevant groups. * indicates a significant difference between the two bracketed conditions. The bottom graphs depict the individual data where AUC values for the EFT condition are plotted against the AUC values for the control condition for the personally relevant group (left graph) and the non-personally relevant group (right graph). Each plotted circle represents one individual. Plotted circles above the dashed line represent participants with higher AUCs in the control condition, whereas plotted circles below the dashed line represent participants with higher AUCs in EFT condition.

Figure 5.12 (top graphs) shows the group AUC values for the EFT and control conditions for each group. A three-way mixed ANOVA (EFT x group x order) showed a significant main effect of EFT, F (1, 122) = 9.96, p = 0.002, and no significant main effects of group, F (1, 122) = 2.10, $p = 0.149$, or order, F (1, 122) = 1.98, $p = 0.162$. There was a significant two-way interaction between EFT and group, F (1, 122) = 6.44, $p = 0.012$. All other interactions were non-significant (all ps > 0.20). Post hoc tests revealed that AUC was significantly higher in the EFT condition ($M = 0.521$, $SD = 0.157$) than in the control condition ($M = 0.463$, $SD = 0.178$) for the personally relevant group ($p < 0.001$). However, there was no significant difference in AUC between the EFT ($M = 0.524$, $SD = 0.175$) and control ($M = 0.517$, $SD = 0.197$) conditions for the non-personally relevant group ($p = 0.642$).

The left bottom graph of Figure 5.12 depicts the individual data for the personally relevant group. Most participants in the personally relevant group (61%) had higher AUCs in the EFT condition compared to the control condition, and 14% of participants had similar AUCs in each condition (i.e., had points that overlapped with the reference line on the graph). The right bottom graph in Figure 5.12 shows the AUCs for the EFT condition compared to the AUCs for control condition for the non-personally relevant group. Fewer than half of the participants in this group (45%) had higher AUCs in the EFT condition compared to in the control condition, and 16% had similar AUCs in each condition. These findings suggest that the personal relevance component of EFT is necessary to reduce delay discounting.

The size of the EFT effect as a function of delay

There was a significant main effect of delay on the size of the EFT effect (log [EFT indifference point/control indifference point]) in the personally relevant group ($X = 9.69$, $p = 0.021$). As shown in Figure 5.13, EFT only reduced discounting at the two-year delay. A one-sample Wilcoxon signed rank test confirmed that the size of the EFT effect was significantly different from zero at the two-year delay ($p < 0.001$).

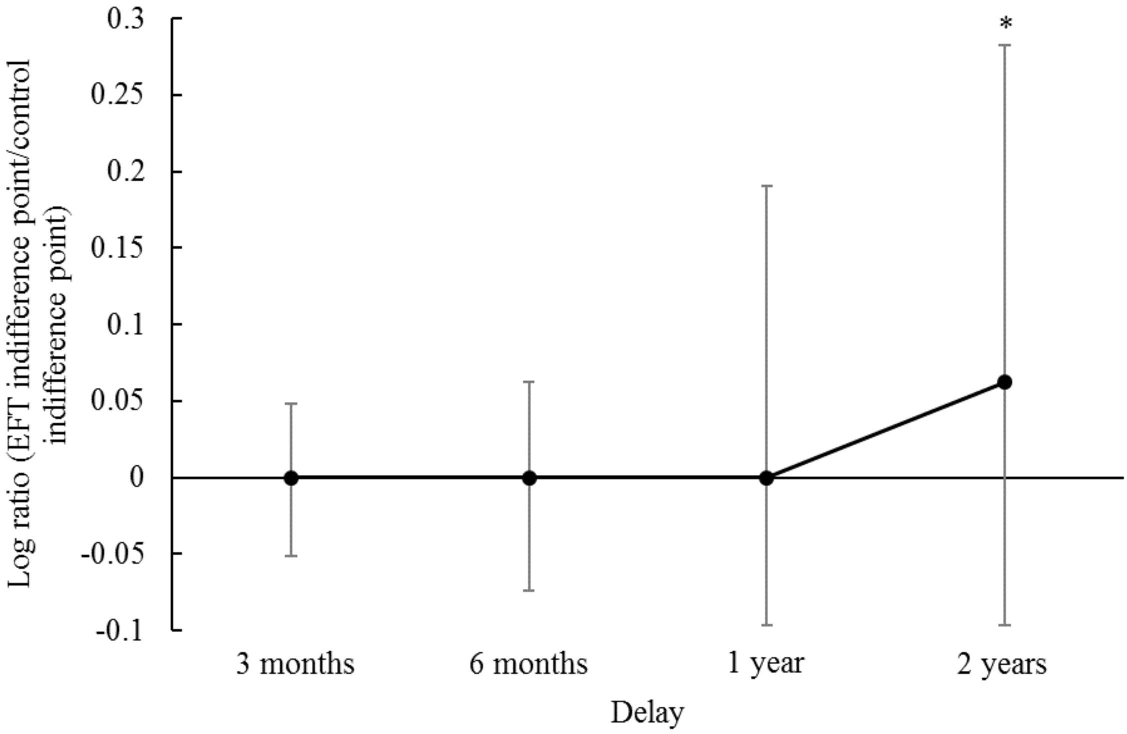

Figure 5.13. The group median size of the EFT effect (log [EFT indifference point/control indifference point]) as a function of delay in the personally relevant group. Points above the x-axis indicate that participants were more self-controlled during the EFT condition than during the control condition. * indicates that the size of the EFT effect was significantly different from zero. The error bars represent the interquartile range.

Manipulation check questions

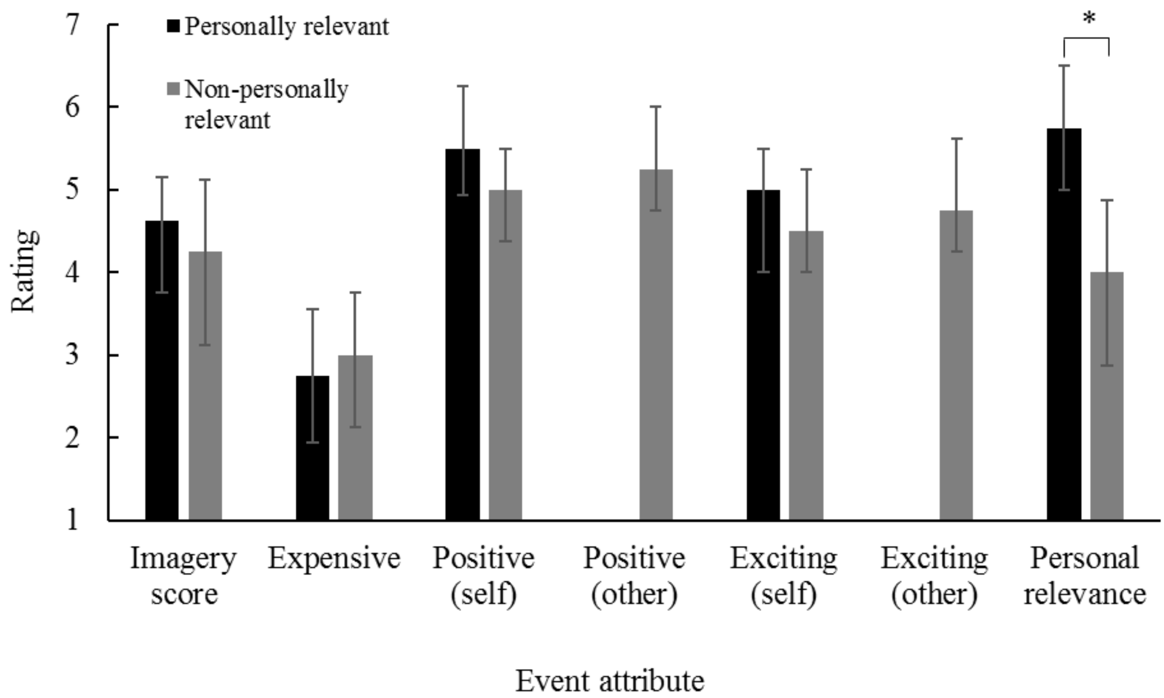

Figure 5.14. Group median ratings for how exciting, expensive, positive and personally relevant the events were for the personally relevant (black bars) and non-personally relevant (grey bars) groups. The group median imagery score (averaged frequency and vividness of imagined events during the money task) is also shown for each condition. * indicates a significant difference between the two bracketed conditions. Error bars represent the interquartile range.

Ratings. Figure 5.14 shows how exciting, expensive, positive and personally relevant the events were rated as for each group, as well as the group median imagery scores. Overall, ratings were high across each event attribute, except for the expensiveness ratings, which were below the midpoint of the scale for each group. The personal relevance ratings were also not particularly high for the non-personally relevant group, as the group median was at the midpoint of the scale. Participants in the personally relevant group rated the events as significantly more personally relevant than participants in the non-personally relevant group ($U = 640.00$, $p < 0.001$), indicating that the personal relevance manipulation was successful. All other differences were non-significant (see Table 5.5 for all p values). There was also no significant difference in the percentage of scene ratings between the two groups at the sequential Bonferroni-corrected alpha level (Personally relevant group: *Mdn* = 50%; Non-personally relevant group: *Mdn* = 70%; $U = 1644.00$, $p = 0.049$), and the medians trend in the opposite direction of the AUC results.

Table 5.5
p values from Mann Whitney U tests comparing event attribute ratings between the personally relevant and non-personally relevant groups

Event attribute	*p* value
Imagery score	0.192
Expensive	0.450
Positive (self)	0.011
Positive (other)	0.483ǂ
Exciting (self)	0.031
Exciting (other)	0.430ǂ
Personal relevance	0.000*

* Significant at the Bonferroni-corrected alpha level.
ǂ These *p* values were obtained by comparing the 'other' ratings from the non-personally relevant group to the 'self' ratings from the personally relevant group.

In contrast to Experiment 3A, there were no significant differences in any of the event ratings between the indifference points with the largest EFT effect and the indifference points with the smallest EFT effect (all *p*s > 0.05). There was also no significant difference in any of the event ratings between EFT indifference points that positively deviated the most from the hyperbolic model and EFT indifference points that deviated the least from the hyperbolic model (all *p*s > 0.05). This non-significant difference is not surprising given that few participants (11%) had unsystematic discounting in the EFT condition according to criterion 1.

'What were you thinking about' questions. The majority of the participants in both the personally relevant (71%) and non-personally relevant (64%) groups stated that they thought about the events during the EFT discounting task, and a Chi-square test showed that there was no significant difference in the frequency with which participants endorsed these statements between the two groups (X^2 (1) = 1.03, $p = 0.310$). Few participants in the personally relevant (5%) and non-personally relevant (1%) groups stated that they thought about events during the control discounting task, and there was no significant difference in frequency between the groups (Fisher's Exact Test $p = 0.326$). These findings indicate that

EFT only occurred during the EFT discounting task for both groups, and to the same extent (consistent with the percent scene ratings).

Forty-nine percent of participants in the personally relevant group explicitly stated that they thought about the financial cost of the events when making decisions during the EFT discounting task, whereas 30% of participants in the non-personally relevant group made these statements. A Chi-Square test revealed that the odds of participants making these statements was 2.3 times higher in the personally relevant group than in the non-personally relevant group, X^2 (1) = 4.72, p = 0.030, *OR* = 2.3. It is not surprising that participants were more likely to think about the cost of the events if they were personally involved in the events. However, one study has shown that, while financial relevance amplifies the EFT effect, it is not necessary to produce the effect (O'Donnell et al., 2017). Therefore, the finding that participants were less likely to think about the financial costs of the events is an unlikely explanation for why there was no EFT effect in the non-personally relevant group.

Thirteen percent of participants in the non-personally relevant group made statements suggesting that they made their decisions during the EFT discounting task based on what their imagined person would do. Such participants made statements such as: "I thought about how they would value the money given the event" or "how much waiting would impact them." It is therefore possible that participants in the non-personally relevant group perceived the EFT discounting task as a task where they were asked to make decisions for their imagined person, rather than for themselves as in the control discounting task. That is, the EFT cues might have prompted participants to decide for the other person, rather than to engage in (non-personal) EFT prior to making a decision for themselves. This could explain the non-significant EFT effect in this group, as people tend to discount rewards for close others (which is perhaps analogous to the non-personal EFT condition) similarly to how they discount rewards for themselves (which is analogous to the control condition) (O'Connell et al., 2013; Ziegler & Tunney, 2012). However, only a small percentage of the participants' statements suggested that they might have interpreted the task in this way, and removal of these participants did not change the results (EFT AUC: M = 0.53, SD = 0.18; Control AUC: M = 0.52, SD = 0.20; t = 0.71, p = 0.478). Figure 5.15 shows the individual data where the grey squares represent the participants who may have discounted for others (n = 9) and the black circles represent the participants who did not indicate that they discounted for others (n = 58).

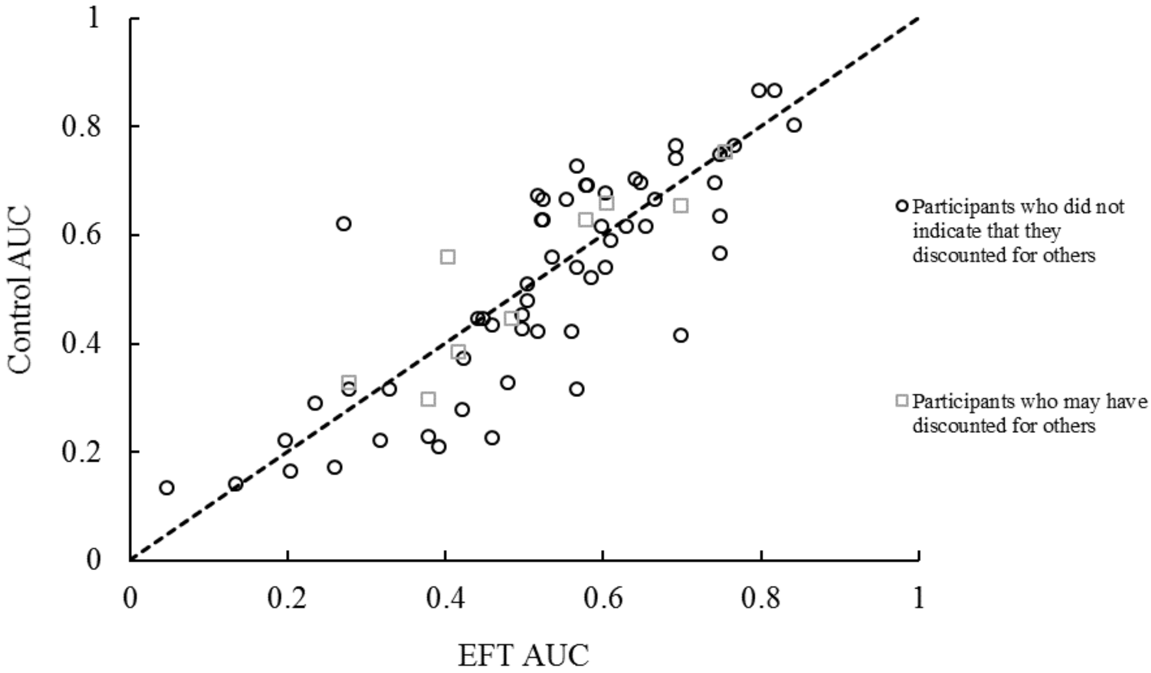

Figure 5.15. AUC values for the EFT condition plotted against the AUC values for the control condition for the non-personally relevant group. Plotted points above the dashed line represent participants with higher AUCs in the control condition, whereas plotted points below the dashed line represent participants with higher AUCs in EFT condition. The grey squares represent participants who indicated that they may have made decisions for the imagined person during the money task (i.e., they 'discounted for others'), whereas the black circles represent participants who did not indicate this.

Demand characteristics

Table 5.6
Participants' answers to the questions assessing demand characteristics coded into categories

	Percentage of participants	
Type of coded answer	Personally relevant group	Non-personally relevant group
Guessed that the researcher was interested in how people make decisions about money, with no mention of the cues (or vice versa)	7% (4/55)	12% (8/67)
Indicated that the researcher predicted that EFT will increase discounting	18% (10/55)	7% (5/67)
Indicated that the researcher predicted that EFT will have no effect on discounting	2% (1/55)	1% (1/67)
Indicated that the researcher predicted that EFT will reduce discounting	36% (20/55)	28% (19/67)
Correctly identified research question, but did not clarify direction of prediction (e.g., investigated whether EFT will affect decisions about money)	27% (15/55)	30% (20/67)
Stated that they did not know the researcher's predictions	7% (4/55)	19% (13/67)
Other ("people don't care about money if happy" or "money distracts people from thinking about people they care about").	2% (1/55)	1% (1/67)

Note: Although more general categories such as "incorrectly guessed the researcher's prediction and results" are shown in Table 5.3 for Experiment 3A, a more specific breakdown of the categories is shown in the current Table 5.6 to allow a clearer comparison between the two groups (given that each group had different AUC results).

As shown in Table 5.6, only 36% of participants in the personally relevant group indicated that the researcher expected EFT to reduce discounting. A further 27% of participants in this group correctly indicated that EFT would affect discounting, but did not specify the direction of the hypothesis. If we are conservative and combine these two categories, then this would mean that 63% of the participants did or could have predicted the results. However, these percentages are similar to that of the non-personally relevant group: 58% of these participants also did (28%) or could have (30%) predicted that the researcher expected EFT to reduce discounting. A Chi-square test confirmed no significant difference in these percentages between the two groups, X^2 (2) = 0.90, p = 0.637.

To further investigate whether demand characteristics can explain the different AUC results between the personally relevant and non-personally relevant groups, we calculated the percentage of participants who predicted that the researcher expected EFT to reduce discounting and also had data that was consistent with this prediction. Figure 5.16 shows these percentages for each group.

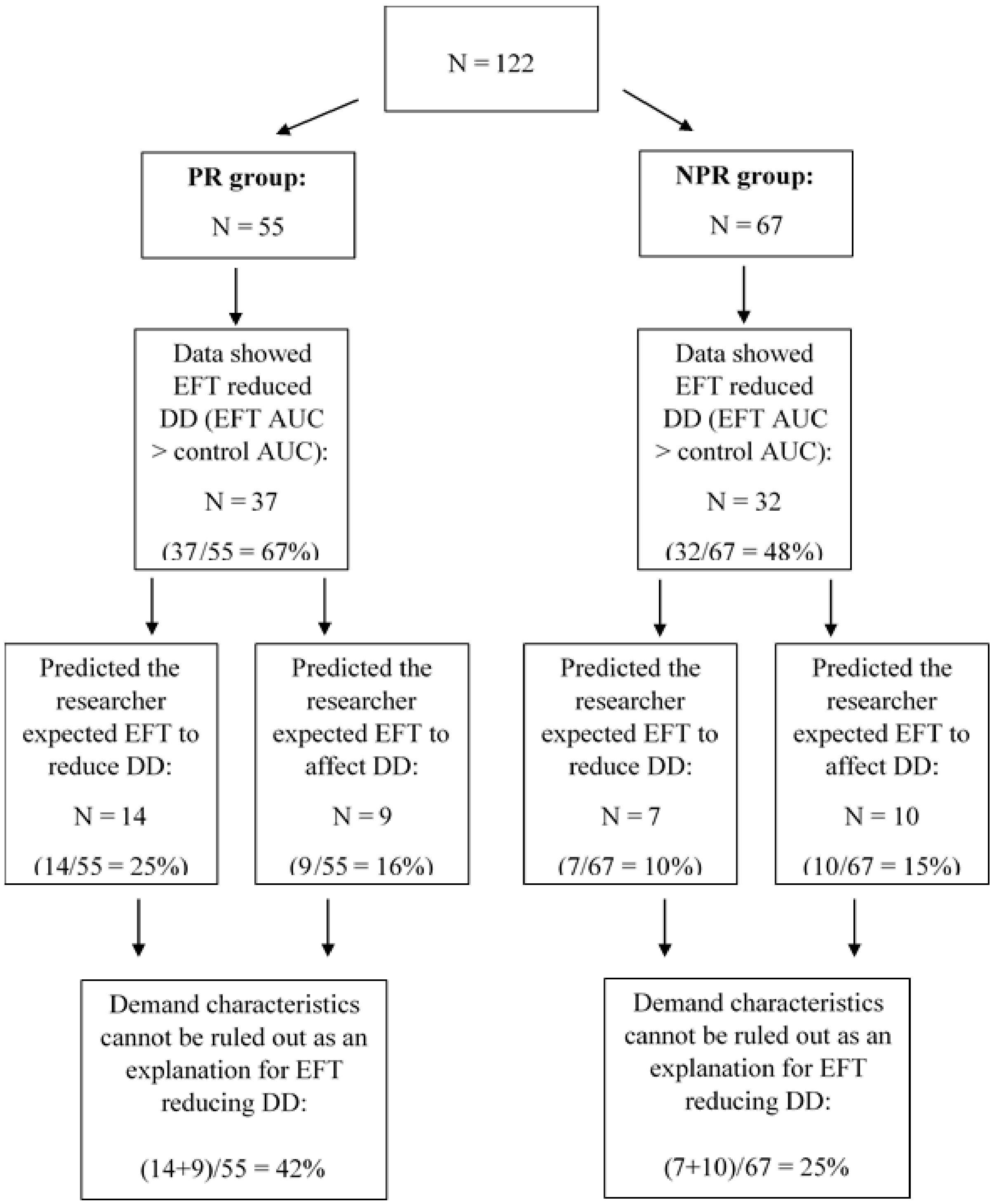

Figure 5.16. Diagram showing the percentage of participants from the personally relevant (PR) and non-personally relevant (NPR) groups who had data showing that EFT reduced their delay discounting (DD) and also predicted (or could have predicted) that the researcher expected this finding.

As shown in Figure 5.16, a higher percentage of participants had data consistent with a demand characteristics account in the personally relevant group than in the non-personally relevant group. However, a Chi-square test showed that these percentages were not significantly different between the two groups (χ^2 (2) = 5.21, p = 0.080). This finding weighs against a demand characteristics account for the different AUC results found between the two groups.

Perceived EFT mechanisms

Table 5.7 shows the percentage of participants from the personally relevant group who selected each option from the 'perceived EFT mechanisms' multiple-choice question. Only participants that had previously correctly indicated that EFT reduced their discounting are included. Participants from the non-personally relevant group were excluded from this analysis given that the current AUC results indicated that non-personally relevant EFT has no significant effect on discounting. Thus, this analysis is based on 24 participants. Most participants stated that EFT reduced their discounting either because: 1) EFT made the future reward seem more valuable, or 2) EFT made the future reward seem more valuable *and* made the delay to the larger reward seem shorter. Few participants stated that EFT reduced discounting by only making the delay to the larger reward seem shorter, or that they did not know why EFT reduced their discounting.

Table 5.7
Percentage of participants from the personally relevant group who stated that EFT reduced their delay discounting because: a) EFT reduced the delay to the larger, later reward, b) EFT increased the value of the larger, later reward, or c) both a and b. The percentage of participants who answered "I don't know why" is also shown. Only participants that had previously correctly indicated that EFT reduced their discounting are included.

Perceived reason for EFT effect	Percentage of participants
EFT reduced the perceived delay to the LL	4%
EFT increased the perceived value of the LL	42%
EFT reduced the perceived delay *and* increased the perceived value of the LL	46%

Perceived reason for EFT effect	Percentage of participants
I don't know why	8%

Note: LL = larger, later reward.

There was no significant difference in the size of the EFT effect (log [EFT AUC/control AUC]) for participants who stated that EFT worked by both reducing perceived delay and increasing perceived value (*Mdn* = 0.04, *N* = 11) than for participants who stated that EFT worked by only increasing perceived value (*Mdn* = 0.11, *N* = 10), *U* = 43.00, *p* = 0.398. However, sample sizes were limited for this analysis.

Overall, these findings suggest that EFT may be conceptualised as an establishing operation that primarily reduces delay discounting by increasing the subjective value of the larger, delayed reward. EFT may also reduce discounting by eliminating the abolishing operation of delay in some people by mentally bringing the future reward into the present (in addition to increasing subjective value). These results also suggest that people may perceive value as more changeable in subjective size than delay.

Conclusion

The current Experiment 3B investigated: 1) whether personal relevance is necessary to reduce delay discounting, and 2) participants' perceptions of the mechanism of the EFT effect. EFT significantly reduced discounting in the personally relevant group, however, there was no significant effect of EFT in the non-personally relevant group. These results show that personal relevance is a necessary component of the EFT effect. Specifically, the future events need to be *highly* personally relevant, given that the non-personally relevant group still rated the events as somewhat personally relevant. We also showed that demand characteristics are an unlikely explanation for these results. These results also cannot be explained by a higher degree of imagery, positivity, excitingness, expensiveness or episodic thinking in the personally relevant group. We found that most participants perceived EFT to reduce their discounting by making the larger, delayed reward seem more valuable, with some participants additionally stating that EFT reduced their discounting by making the delay to the larger reward seem shorter. Whether one or both of these mechanisms were mentioned did not significantly affect the size of the EFT effect. However, we could only conduct this analysis on a limited number of individuals.

Chapter 6: General Discussion

The current thesis examined the reliability of the effect of EFT on delay discounting and also sought to clarify the components of EFT that are necessary to reduce discounting. Experiment 1 replicated the EFT effect using a common variant of the two choice, titrating-amount procedure, and the meta-analysis confirmed that EFT has a reliable, medium-sized effect on delay discounting. Experiments 2A, 2B and 2C assessed the suitability of an alternative and briefer report indifference points procedure for subsequent EFT experiments, and found that this procedure does not provide a sufficiently valid measure of discounting. Experiment 3A found that episodic and future thinking are necessary components of the EFT effect, and that financial relevance alone is not sufficient to reducing discounting. Experiment 3B found that the future events also need to be personally relevant to reduce discounting. We also showed that demand characteristics are an unlikely explanation for the EFT effects found in Experiments 3A and 3B. Rather, EFT primarily reduced discounting by increasing the perceived value of the larger, delayed reward. Table 6.1 shows a summary of the size of the EFT effect (Hedges' *g*) and other relevant findings across Experiments 1, 3A and 3B. Below we discuss potential reasons for the variation in the size of the EFT effect across studies, as well as other inconsistencies. We also consider whether the current findings are consistent with previously proposed mechanisms of the EFT effect. Finally, we discuss the therapeutic potential of EFT within the context of the current findings.

Table 6.1
Summary of effect sizes (Hedges' g), percentage of participants with unsystematic discounting in the EFT condition, group median event ratings and delays with EFT effects across EFT studies of the current thesis.

Study	Hedges' *g*	Percent unsystematic in EFT	Positivity rating	Excitingness rating	Expensiveness rating	Imagery score	Delays with EFT effect
E1	0.614	28% (flat function = 15%)	6	5.5	2	4.5	1 month-1 year
E3A	0.535*	39% (flat function = 2%)	6	6	3.83	5	6 months-1 year

Study	Hedges' *g*	Percent unsystematic in EFT	Positivity rating	Excitingness rating	Expensiveness rating	Imagery score	Delays with EFT effect
E3B	0.343	12% (flat function = 0%)	5.5	5	2.75	4.63	2 years

Note: Hedges' *g* was calculated using Equations 3.9 and 3.12 (see Chapter 3 [meta-analysis]).
* This effect size is the average of the three effect sizes that were calculated using the NT, SFT and EPT control conditions.

Event characteristics as an explanation for differences in EFT effect sizes across studies

It is possible that the size of the EFT effect was smallest in Experiment 3B because the events participants imagined were less positive and exciting than the events in Experiments 1 and 3A, as shown in Table 6.1. Kruskal Wallis tests confirmed that the positivity, excitingness and expensiveness ratings were significantly different across experiments (all *p*s < 0.002), but imagery scores were not ($p = 0.143$). Follow-up Mann Whitney U tests showed that the excitingness and positivity ratings were significantly lower in Experiment 3B than in Experiments 1 and 3A (*p*s < 0.008), likely because participants did not generate their own future events. Expensiveness ratings were significantly higher in Experiment 3A than in Experiments 1 and 3B (*p*s < 0.002). All other rating differences were non-significant (*p*s > 0.5). However, these analyses are exploratory, given that participants were not randomly assigned to each experiment.

To our knowledge, no study has assessed whether degree of positivity or excitingness of events is related to the size of the EFT effect, as most researchers have ensured that the different events were matched on these dimensions. We therefore determined whether the size of the EFT effect was correlated with the excitingness and positivity ratings within Experiments 1, 3A and 3B. The scatter plots are shown in Figure 6.1.

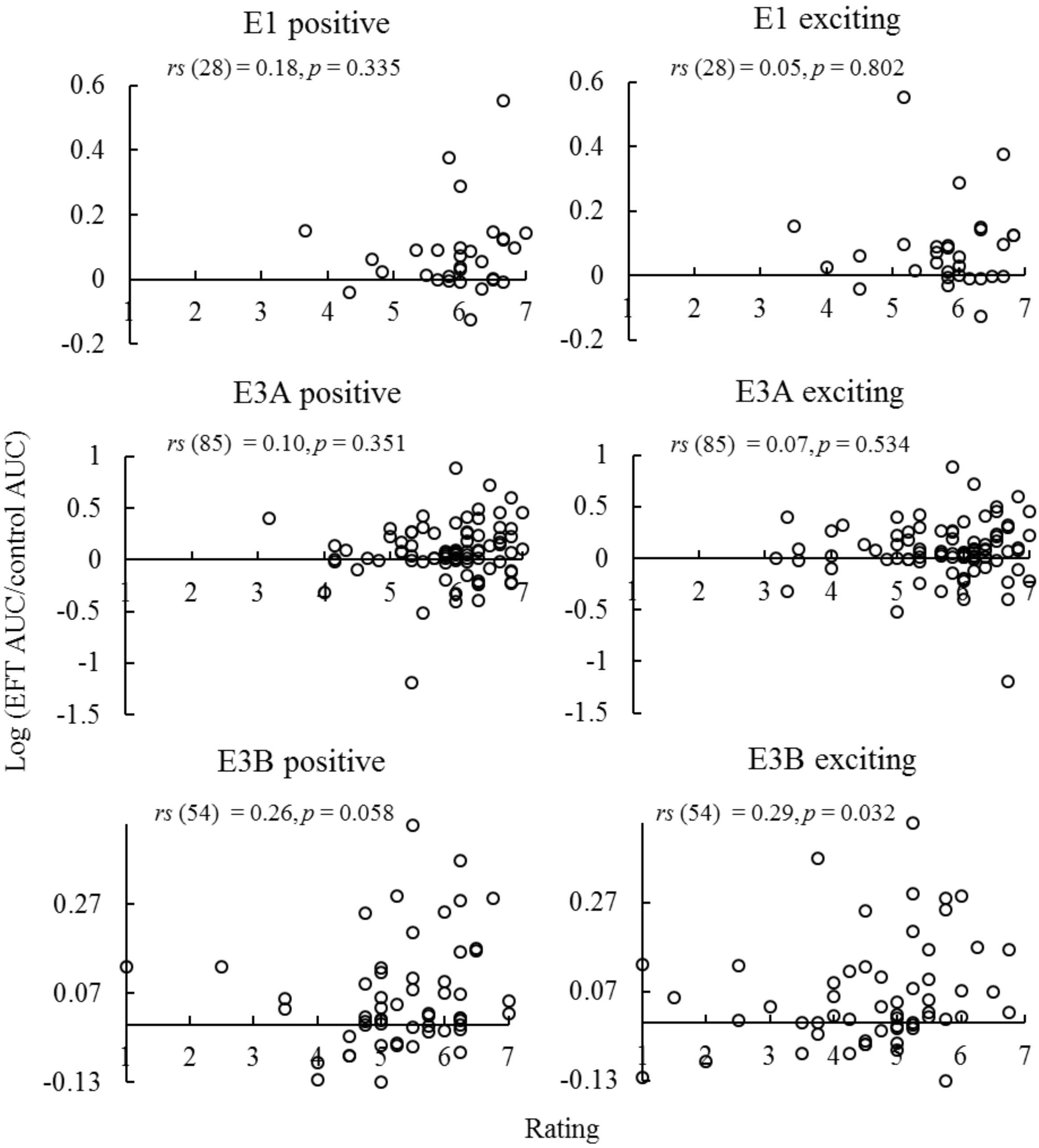

Figure 6.1. The size of the EFT effect (log [EFT AUC/control AUC]) as a function of event ratings (averaged across delays) for Experiments 1 (top graphs), 3A (middle graphs) and 3B (bottom graphs). The left graphs are for the positivity ratings and the right graphs are for the excitingness ratings. Points above the x-axis are participants who were more self-controlled during the EFT condition than during the control condition.

As shown in Figure 6.1, the size of the EFT effect was significantly positively correlated with the excitingness ratings, but only in Experiment 3B. The non-significant correlations in Experiments 1 and 3A may be due to the fact that most participants rated the events as highly exciting (and positive) in those studies; few participants gave ratings of less than 4. In contrast, there was a higher degree of variability in the excitingness ratings for

Experiment 3B, as more participants also gave ratings of less than 4. However, the significant correlation in Experiment 3B is weak, and there were still fewer participants who rated the events they thought about on the lower end of the scale overall. Greater variability in the event ratings (i.e., more participants with lower scores) would help to clarify whether the size of the EFT effect is related to how exciting and/or positive the future events are. In a future study, participants could be instructed to write about a range of positive/exciting future events at each delay and to rate them from least positive/exciting to most positive/exciting. We could then determine whether the size of the EFT effect is greater when participants are cued to think about the most positive/exciting future events than when they are cued to think about the least positive/exciting future events.

The possible relationship between degree of positivity/excitingness and the size of the EFT effect is consistent with previous studies showing that positive emotional valence is an important contributor to the EFT effect. Overall, research to date indicates that positively-valenced EFT is likely to be most effective at reducing discounting, given that negative and neutral EFT have produced inconsistent results (Bulley et al., 2019; Lin & Epstein, 2014; Lempert et al., 2012; Liu et al., 2013, Zhang et al., 2018; see meta-analysis). A recently published study by Calluso, Tosoni, Cannito and Committeri (2019) further supports this conclusion. In this study, participants completed a standard delay discounting task without any cues (i.e., a NT control), followed by a discounting task where they were cued to imagine positive, negative and neutral future events. The positive, negative and neutral cues were blocked in a counterbalanced order. Positive, negative and neutral EFT all significantly reduced discounting compared to the NT control. However, the size of the EFT effect was larger for positive EFT than for neutral EFT, and smallest for negative EFT.

Unsystematic discounting is a consequence of the EFT manipulation

The percentage of participants with unsystematic discounting was also notably lower in Experiment 3B than in Experiments 1 and 3A, which is consistent with the conclusion that unsystematic discounting is a consequence of an effective EFT manipulation, given that the size of the EFT effect was smallest in Experiment 3B. To reiterate, an extremely effective EFT manipulation might cause people to continuously choose the larger, delayed reward (i.e., produce a flat discounting function), which would violate Johnson and Bickel's second (2008) criterion for systematic discounting. Consistent with this possibility, Experiment 1 had the largest percentage of flat functions and the largest EFT effect. An effective EFT manipulation might also produce a discounting function that violates Johnson and Bickel's first criterion (i.e., where an indifference point is greater than the previous by more than 20% of the larger, delayed reward) if the different events are not matched on particular attributes. As a result, some events might produce a larger EFT effect for an individual than others.

To investigate the latter possibility, we identified the event ratings at the delay where the EFT indifference point positively deviated the most from the hyperbola-predicted indifference point (i.e., the highest log ratio [log {EFT indifference point/hyperbola-predicted EFT indifference point}]) and at the delay where the EFT indifference point deviated the least from the hyperbola-predicted EFT indifference point (i.e., the smallest log ratio) for each individual. We compared the ratings at these two delays given that an indifference point that is deemed unsystematic is more likely to deviate from the hyperbola than an indifference point that is deemed systematic. We found that, in Experiment 3A, EFT indifference points that positively deviated the most from the hyperbolic model prediction had significantly higher excitingness ratings than EFT indifference points that deviated the least from the hyperbolic model. This finding was specific to Experiment 3A, which had the highest percentage of participants with unsystematic discounting (with few of these being characterised as a flat discounting function). Thus, the high percentage of unsystematic discounting in Experiment 3A might be explained by the fact that the events were not matched in excitingness.

The EFT indifference points that positively deviated the most from the hyperbolic model also had significantly higher expensiveness ratings than EFT indifference points that deviated the least from the hyperbolic model at the 0.05 alpha level in Experiment 3A. This indicates that the unsystematic discounting might also be explained by the fact that the events were not matched in expensiveness. Consistent with this explanation, expensiveness ratings were also significantly higher at the indifference points with the largest EFT effect (i.e., the highest log ratio of the EFT and control indifference points [log {EFT indifference point/control indifference point}]) than at the indifference points with the smallest EFT effect (i.e., the smallest log ratio of the EFT and control indifference points). Fewer participants had unsystematic discounting according to criterion 1 in Experiments 1 and 3B, indicating that the events were rated more similarly across the delays. Specifically, approximately half of the participants with unsystematic discounting in Experiment 1 were deemed unsystematic because they continuously chose the larger, delayed reward. In Experiment 3B, a low percentage of participants showed unsystematic discounting overall, likely because the different events were more standardised across the different delays, and were rated as less positive and exciting overall, as noted above.

Additionally, the EFT effect was either weaker or rendered non-significant when unsystematic participants were excluded from analyses in Experiments 1 and 3A. These findings further support the conclusion that the participants with unsystematic discounting were most affected by the EFT manipulation. Thus, as previously discussed in the meta-analysis, Johnson and Bickel's (2008) criteria to identify unsystematic discounting should not be applied to an EFT condition with the goal of identifying participant inattentiveness or misunderstanding of the task. Rather, researchers should apply these criteria to a NT control condition only or implement attention check questions to ensure data quality (e.g., see Sze et al., 2017). However, if researchers ensure that the different events are matched on various attributes (e.g., as in Peters & Buchel, 2010), then Johnson and Bickel's first criterion could be applied to EFT conditions in order to identify random responding.

Further discussion of the hyperbolic model and discounting measures

Given the high levels of unsystematic discounting in Experiments 1 and 3A, it is not surprising that the hyperbolic model did not describe discounting well in the EFT conditions. Even after participants with unsystematic discounting had been removed, the hyperbolic model still did not describe these data well, as a substantial proportion of the remaining participants had data that approached, but did not quite meet, Johnson and Bickel's (2008) criteria for unsystematic discounting. Inspection of these data did not suggest a systematic (but non-hyperbolic) discounting function, indicating that the data would not be better described by another theoretical model. Although there were low levels of unsystematic discounting in Experiment 3B overall, a substantial proportion of participants still had data that were not well described by the hyperbolic model. Some of these participants in the EFT condition approached, but did not quite meet, criteria for unsystematic discounting, as in Experiments 1 and 3A. Thus, the poor model fit in Experiment 3B may also be a consequence of the EFT manipulation. However, the EFT effect was not as strong in Experiment 3B compared to Experiments 1 and 3A where more participants did meet criteria for unsystematic discounting. In sum, an effective EFT manipulation may not produce hyperbolic discounting. Rather, EFT may produce an unsystematic discounting function. EFT researchers should therefore utilise discounting measures that do not require any assumptions regarding the mathematical form of the discounting function (such as AUC).

The poor hyperbolic model fit in Experiment 3B may also be attributed to the range of delays that were used. In contrast to Experiments 1 and 3A, Experiment 3B used longer delays. Given the general pattern that reinforcer value decreases less as delays get longer, indifference points were more similar to each other in Experiment 3B than in other experiments using a wider range of delays. Given the indifference points typically varied less, each indifference point did not vary as much from the mean of all indifferent points for each individual. A small deviation in the indifference points from the mean would produce a low R^2, given that R^2 is calculated by dividing the sum of squares error (SSE) by the sum of squared deviations of the indifference points from the mean indifference point (SSM) and subtracting this from 1. Indeed, the mean SSM was smaller in Experiment 3B (M = 0.03) than in Experiments 1 (M = 0.04) and 3A (M = 0.05).

Although the poor model fits in Experiments 1, 3A and 3B may be explained as a consequence of the EFT manipulation or the delay range, the poor model fits in the report indifference points procedure (Experiments 2A, 2B and 2C) was due to difficulties participants had with the task. That is, participants were unable to accurately report their own indifference points, and therefore the report indifference points procedure did not provide a valid measure of discounting.

Another major finding that has been discussed throughout the current thesis is that EFT effects were larger when AUC was used as a discounting measure compared to k (see meta-analysis and Experiment 3A). We suggested that AUC might produce larger EFT effects because longer delays contribute more to the AUC than shorter delays (Borges et al., 2016), and EFT has been shown to produce larger effects at longer delays (Snider et al., 2016; Stein et al., 2016). Consistent with this explanation, Experiments 1, 3A and 3B found that an EFT effect was only present at the longest delays. However, an inconsistency across these experiments was the absolute delay-length that showed an EFT effect. As shown in Table 6.1, Experiments 1 and 3A both showed EFT effects at delays of six months and one year. However, the EFT effect was not significant at these delays in Experiment 3B; EFT only reduced discounting at the longest delay of two years. It is likely that EFT produces larger effects at longer delays than at shorter delays because the effect of discounting is greatest at the longer delays and therefore a greater EFT-induced change is possible compared to baseline. Weaker EFT effects (i.e., as in Experiment 3B relative to Experiments 1 and 3A) are therefore less likely to affect discounting at shorter delays.

However, EFT might be less effective at extremely long delays, given that people find it easier to imagine events in the near future than in the far future (Arnold, McDermott & Szpunar, 2011; D'Argembeau, Renaud, & Van der Linden, 2011; D'Argembeau & Van der Linden, 2004). D'Argembeau and Van der Linden (2004) found that closer future events (1 month-1 year) contained significantly more sensorial/contextual details and stronger feelings of pre-experiencing than relatively distant future events (5-10 years). EFT may therefore become less effective at delays as long as five years (or longer than two years). Future research is needed to determine the future time point for which EFT may become ineffective. Mok (2017) reported that indifference points were significantly higher in an EFT condition compared to a control condition at a five-year delay. Few studies have examined EFT at longer delays (see meta-analysis), and those that have did not report whether EFT reduced indifference points at these longer delays.

Further discussion of demand characteristics

We found that demand characteristics were an unlikely explanation for the results in both Experiments 3A and 3B. However, the design of our experiments may have made it harder for participants to guess the research question, compared to other EFT studies. For example, in Experiment 3A we had multiple conditions that involved the use of cues (not just EFT) manipulated within subjects. We also did not clearly associate the cues with the larger, delayed reward during the money task in Experiments 3A and 3B (as in other EFT studies, such as Experiment 1 of the current thesis where the cue was written below the larger, delayed option). Rather, the cue word was positioned between the two monetary options in Experiments 3A and 3B. Comparison of the effect sizes among Experiments 1, 3A and 3B support the theory that these methodological differences could have affected the likelihood of demand characteristics, as the size of the EFT effect was largest in Experiment 1.

A recently published study by Rung and Madden (2019) attempted to separate the theoretically relevant and demand-characteristic components of EFT. All participants were prompted to imagine future events at specific time points prior to completing a money-discounting task. Participants in an 'EFT-event' group had cues during the money task that referred to the event but not the delay (e.g., "imagine you are at the beach", as in experiments of the current thesis), whereas participants in an 'EFT-time' group had cues that referred to the delay but not the event (e.g., "imagine 1 year from now"). The 'EFT-event' cues were intended to be high in theoretical relevance (i.e., episodic thinking) but low in demand characteristics, whereas the 'EFT-time' cues were intended to be low in theoretical relevance but high in demand characteristics (given that the matching delays between EFT and the money task were emphasised). EFT only reduced discounting in the 'EFT-event' group, however, only participants in the 'EFT-event' group were able to correctly identify the hypothesis at greater than chance levels. Thus, the cues with high theoretical relevance were actually high in demand characteristics, rather than the cues that were low in theoretical relevance. Nevertheless, reduced discounting was unrelated to whether participants identified the hypothesis, indicating that demand characteristics did not contribute to the EFT effect. These results also indicate that episodically thinking about the future event is necessary to reduce discounting, consistent with Experiment 3A.

Further discussion of the theoretical mechanism(s) of the EFT effect

We previously suggested that EFT may reduce delay discounting by: 1) increasing the subjective value of the larger, delayed reward by making the larger, delayed reward more concrete (Kim et al., 2013; Trope & Liberman, 2003; Yi et al., 2017), 2) broadening an individual's time frame over which they consider the consequences of their behaviour (i.e., broadening "temporal horizon"; Lin & Epstein, 2014; Snider et al., 2016), and/or 3) decreasing the perceived delay to the larger, delayed reward. Findings across the current thesis are most consistent with the first proposed mechanism. Firstly, our finding that non-personally relevant EFT had no effect on discounting is consistent with construal level theory, from which the first proposed mechanism is derived. To reiterate, construal level theory states that not only does the construal of information depend on temporal distance, but also on social distance. That is, the more similar someone is to us, the more socially close they seem and are therefore construed more concretely. In contrast, the more dissimilar someone is to us, the more abstractly they will be construed (Trope & Liberman, 2003; Liviatan et al., 2008). Thus, thinking about someone else's future events will not make the larger, delayed reward as concrete as thinking about your own future events will. Construal level theory therefore predicts that non-personal EFT is less likely to reduce discounting than personal EFT.

Secondly, most participants stated that EFT primarily reduced their discounting by increasing the perceived value of the larger, delayed reward, rather than by reducing the perceived delay to the larger, delayed reward. This further supports the theory that EFT may be conceptualized as an establishing operation (an event that alters the value of a consequence). Establishing operations have temporary effects on behaviour (Pierce & Chaney, 2017), which is also consistent with findings from Experiment 3A which showed that EFT does not produce long-lasting effects on discounting. However, there were still some individuals who stated that EFT also reduced their discounting by making the delay to the larger reward seem shorter. Additionally, this analysis was based on a limited number of participants. Future research should further investigate whether EFT reduces discounting by altering time perception.

One commonly used time perception task is the temporal bisection task, where participants are required to state whether a stimulus was presented for a relatively short or long period of time, for example, two seconds versus four seconds (Church & Deluty, 1977; also see Kopec & Brody, 2010 for a review). Prior to beginning the task, participants are typically trained using two reference stimuli that are labelled as short or long. Thus, the participant must indicate which reference stimulus (short or long) they believe each test stimulus is most similar to, and the proportion of long and short choices are used to assess temporal perception. Baumann and Odum (2012) found that steep delay discounting was associated with overestimation of time (as measured by the temporal bisection task). Future research could investigate whether EFT reduces discounting by reducing perceived delay by having participants complete a temporal bisection task, followed by an EFT discounting task, followed by the same temporal bisection task again. Participants in a control group could complete the same tasks without an EFT manipulation. If participants in the EFT condition perceive time as shorter in the second temporal bisection task than in the first temporal bisection task, then this would indicate that EFT reduces subjective time.

However, one concern with using the temporal bisection task is that the time intervals must be relatively short (in seconds or minutes), whereas EFT only reduces discounting at delays as short as one month (Snider et al., 2016; Experiment 1 of the current thesis). Additionally, Baumann and Odum (2012) found that the temporal bisection task was only weakly correlated with discounting rate, possibly due to the different time spans (2-4 seconds vs. 1 week-6 years). Thus, an alternative way to measure time perception could be presenting participants with the written delays corresponding to the EFT delays and asking them to rate them on a scale from short to long (e.g., 1 = short; 7 = long). Participants may need to be prompted to imagine the corresponding event prior to rating each delay at time 2, given that EFT does not produce long-lasting effects (see further discussion of the therapeutic potential of EFT below). Although demand characteristics are perhaps more of a concern using this procedure, this could be addressed by asking participants what they thought the research question and hypotheses were, as in experiments of the current thesis.

We previously suggested that, if EFT reduces discounting by broadening an individual's time frame over which they consider the consequences of their behaviour, then EPT might also reduce discounting. However, we found that EPT had no effect on discounting. Although this finding does not necessarily contradict a broadening time-frame account (as EFT may still broaden how far one thinks into the *future*), one recently published study suggests that this may not be the mechanism through which EFT reduces discounting. Rung and Madden (2019) assessed whether broadened temporal horizons mediate the effect of EFT on discounting. Participants were randomly assigned to an EFT condition or a control (episodic recent thinking) condition. Following the EFT delay discounting task, participants completed the future time perspective task (Wallace, 1956), which is designed to measure how far one thinks into the future. During this task, participants are instructed to list events that are likely to happen throughout their lives and to specify their ages at each event. EFT significantly reduced discounting, and future horizons were significantly shorter in the control group than in the EFT group, indicating that expanded future horizons might be the mechanism through which EFT reduces discounting. However, expanded future horizons did not significantly mediate the effect of EFT on discounting. It is possible that expanded future horizons did not mediate the EFT effect because the EFT cues were not presented during the future time perspective task. However, Rung and Madden noted that this cannot explain why future horizons were significantly shorter in the control group than in the EFT group.

In summary, the current thesis contributes to our current understanding of the workings of EFT by investigating the components of EFT that are necessary to reduce discounting. However, further research is warranted to form a better understanding of the mechanism(s) through which EFT reduces discounting. For the moment, a change in construal is a candidate account, although changes in perceived delay and temporal horizon also warrant further investigation.

The therapeutic potential of EFT

The current thesis provides several translational implications for the use of EFT in clinically relevant settings. Findings from the meta-analysis that the EFT effect is robust across different participant characteristics and ages indicate that EFT may be utilized as an effective intervention for a variety of individuals. Findings from Experiment 3B show that individuals need to focus on personally relevant aspects of the future event. Thirdly, findings from Experiments 3A and 3B indicate that imagining the most exciting future events may be most effective in enhancing self-control. However, fantasy realization theory (Oettingen, 1996; 1999) suggests that individuals should ensure that the events are not so exciting that they become unrealistic, as this could reduce the EFT effect.

It is also important to consider the practical limitations of EFT. Experiment 3A found that EFT does not produce a lasting effect, as participants needed to be continuously cued for EFT to reduce their discounting. We also found no evidence of a carry-over effect of EFT in Experiment 1 when the EFT and control trials were interspersed (as well as no order effects in Experiments 3A and 3B), further showing that EFT does not generalize to control delays when participants are not cued. The fact that the excitingness/expensiveness ratings of the cues may have determined how much a single indifference point deviated from the hyperbolic curve in Experiment 3A further highlights the specificity of the EFT effect and the importance of the cues. These overall findings suggest that a brief course of EFT may be utilised as an effective intervention for helping people make self-controlled, one-off decisions, such as whether to opt into a retirement savings plan.

Mellis, Snider, Deshpande, LaConte and Bickel (2019) have recently published research (after all data for the current thesis were collected) also showing that repeated administration of EFT does not produce a lasting effect on delay discounting. Participants in Mellis et al.’s (2019) study completed six identical sessions. During each session, participants completed a standard delay discounting task (i.e., a NT control), described five future events and completed a discounting task where they were cued to think about the future events. EFT significantly reduced discounting at all but the first session, and the size of the EFT effect increased as a function of session number. Participants did not rate the events as more vivid, important, exciting, or enjoyable as session number increased. The authors therefore concluded that repeated cue exposure increased sensitivity to the cues. Additionally, discounting in the NT control remained stable over time, consistent with the current experiments. Mellis et al.’s study shows that, although repeated administrations of EFT may be most effective for increasing self-control, this still does not have a long lasting effect on self-control, as with one-off EFT administrations. Thus, for EFT to serve as an effective intervention, individuals must be continuously cued to think about the future events.

Research has already demonstrated the translational practicality of continuous cuing. Sze et al. (2015) conducted a small feasibility study where they delivered audio and written EFT cues to participants' smartphones via a web-based system over a four-week period. The participants reported that the system was easy and helpful to use. Additionally, although the study was not powered to detect a meaningful change in body weight, the authors reported a promising trend towards EFT-reduced body weight. A similar approach has been used by O'Neill et al. (2016), who found that audio EFT cues delivered to participants' smartphones significantly reduced calorie intake within a food court setting. Mellis et al. (2019) suggested that EFT could also be an effective intervention without cues if individuals are able to think about the future events out of habit.

Conclusion

The present thesis examined the reliability of the EFT effect both experimentally and meta-analytically, and showed that EFT has a significant, small to medium-sized effect on delay discounting. These results strengthen the conclusion that EFT is a promising intervention that may be used to reduce impulsive decision making, given that delay discounting is a known contributor to a variety of impulsive behaviours. The present thesis also contributes to the literature by investigating the components of EFT that are necessary to reduce delay discounting. Novel findings were that the episodic thinking, future thinking and personal relevance components are all necessary to reduce discounting, and that financial relevance alone is not sufficient. We also found that demand characteristics are an unlikely explanation for these findings. This was an important alternative explanation to address to ensure that the findings can be successfully translated into real-world settings. The present thesis also contributes to the delay discounting literature more generally by examining the validity of the report indifference points procedure and by examining the impact that different discounting measures can have on researchers' results. Continued investigations into the circumstances under which EFT reduces discounting (both within and outside a laboratory setting) will be important for informing how best to translate EFT into clinical settings.

References

References marked with an asterisk indicate studies included in the meta-analysis

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia, 45*, 1363-1377. https://doi.org/10.1016/j.neuropsychologia.2006.10.016

Addis, D. R., Wong, A. T., & Schacter, D. L. (2008). Age-related changes in the episodic simulation of future events. *Psychological Science, 19(1)*, 33-41. https://doi.org/10.1111/j.1467-9280.2008.02043.x

Albrecht, K., Volz, K., Sutter, M., Laibson, D., & von Cramon, Y. (2011). What is for me is not for you: Brain correlates of intertemporal choice for self and other. *Social Cognitive and Affective Neuroscience, 6*(2), 218-225. https://doi.org/10.1093/scan/nsq046

Amlung, M., Petker, T., Jackson, J., Balodis, I., & MacKillop, J. (2016). Steep discounting of delayed monetary and food rewards in obesity: A meta-analysis. *Psychological Medicine, 46*, 2423–2434. https://doi.org/10.1017/S0033291716000866

Arnold, K. M., McDermott, K. B., & Szpunar, K. K. (2011). Imagining the near and far future: The role of location familiarity. *Memory & Cognition, 39*(6), 954-967. https://doi.org/10.3758/s13421-011-0076-1

Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. *Trends in Cognitive Sciences, 5(12)*, 533-539. https://doi.org/10.1016/S1364-6613(00)01804-0

Baker, F., Johnson, M. W., & Bickel, W. K. (2003). Delay discounting in current and never-before cigarette smokers: Similarities and differences across commodity, sign, and magnitude. *Journal of Abnormal Psychology, 112*, 382–392. https://doi.org/10.1037/0021-843X.112.3.382

*Banes, K. E. (2016). *The Effect of Episodic Future Thought on Delay Discounting, Outcome Expectancies, and Alcohol Use among Risky College Drinkers*. Unpublished doctoral dissertation, Virginia Tech.

Bartels, D. M., & Rips, L. J. (2010). Psychological connectedness and intertemporal choice. *Journal of Experimental Psychology: General, 139*, 49-69. https://doi.org/10.1037/a0018062

Baumann, A. A., & Odum, A. L. (2012). Impulsivity, risk taking, and timing. *Behavioural Processes, 90*(3), 408-414. https://doi.org/10.1016/j.beproc.2012.04.005

Bechara, A., & Martin, E. M. (2004). Impaired decision making related to working memory deficits in individuals with substance addictions. *Neuropsychology, 18*, 152–162. https://doi.org/10.1037/0894-4105.18.1.152

*Benoit, R. G., Gilbert, S. J., & Burgess, P. W. (2011). A neural mechanism mediating the impact of episodic prospection on farsighted decisions. *Journal of Neuroscience, 31(18)*, 6771-6779. https://doi.org/10.1523/JNEUROSCI.6559-10.2011

Benzion, U., Rapoport, A., & Yagil, J. (1989). Discount rates inferred from decisions: An experimental study. *Management Science, 35*(3), 270-284. https://doi.org/10.1287/mnsc.35.3.270

Bialaszek, W., Bakun, P., McGoun, E., & Zielonka, P. (2016). Standing in Your Peer's Shoes Hurts Your Feats: The Self-Other Discrepancy in Risk Attitude and Impulsivity. *Frontiers in Psychology*, 7, 197. https://doi.org/10.3389/fpsyg.2016.00197

Bickel, W. K., & Marsch, L. A. (2001). Toward a behavioral economic understanding of drug dependence: delay discounting processes. *Addiction, 96(1)*, 73-86. https://doi.org/10.1046/j.1360-0443.2001.961736.x

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

Boettiger, C. A., Mitchell, J. M., Tavares, V. C., Robertson, M., Joslyn, G., D'Esposito, M., & Fields, H. L. (2007). Immediate reward bias in humans: fronto-parietal networks and a role for the catechol-O-methyltransferase 158Val/Val genotype. *Journal of Neuroscience, 27(52)*, 14383-14391. https://doi.org/10.1523/JNEUROSCI.2551-07.2007

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-analysis*. Wiltshire, England: Wiley & Sons, Ltd. https://doi.org/10.1002/9780470743386

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2013). *Comprehensive meta-analysis version 3*. Englewood, NJ: Biostat, Inc.

Borenstein, M., & Higgins, J. P. (2013). Meta-analysis and subgroups. *Prevention Science, 14(2)*, 134-143. https://doi.org/10.1007/s11121-013-0377-7

Borges, A. M., Kuang, J., Milhorn, H., & Yi, R. (2016). An alternative approach to calculating Area-Under-the-Curve (AUC) in delay discounting research. *Journal of the Experimental Analysis of Behavior, 106*(2), 145-155. https://doi.org/10.1002/jeab.219

Botzung, A., Denkova, E., & Manning, L. (2008). Experiencing past and future personal events: Functional neuroimaging evidence on the neural bases of mental time travel. *Brain and Cognition, 66*, 202-212. https://doi.org/10.1016/j.bandc.2007.07.011

*Bromberg, U., Lobatcheva, M., & Peters, J. (2017). Episodic future thinking reduces temporal discounting in healthy adolescents. *PloS one, 12(11)*, e0188079. https://doi.org/10.1371/journal.pone.0188079

Buckner, R. L. (2004). Memory and executive function in aging and AD: multiple factors that cause decline and reserve factors that compensate. *Neuron, 44(1)*, 195-208. https://doi.org/10.1016/j.neuron.2004.09.006

*Bulley, A., & Gullo, M. J. (2017). The influence of episodic foresight on delay discounting and demand for alcohol. *Addictive Behaviors, 66*, 1-6. https://doi.org/10.1016/j.addbeh.2016.11.003

*Bulley, A., Miloyan, B., Pepper, G. V., Gullo, M. J., Henry, J. D., & Suddendorf, T. (2019). Cuing both positive and negative episodic foresight reduces delay discounting but does not affect risk-taking. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/1747021818819777

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach - Second edition*. New York, New York: Spnnger-Veflag.

Callan, M. J., Shead, N. W., & Olson, J. M. (2011). Personal relative deprivation, delay discounting, and gambling. *Journal of Personality and Social Psychology, 101*, 955-973. https://doi.org/10.1037/a0024778

Calluso, C., Tosoni, A., Cannito, L., & Committeri, G. (2019). Concreteness and emotional valence of episodic future thinking (EFT) independently affect the dynamics of intertemporal decisions. *PloS one, 14*(5), e0217224.

https://doi.org/10.1371/journal.pone.0217224

Cassidy, R. N., & Kangas, B. D. (2014). Impulsive students participate later: Delay discounting in a research subject pool. *Experimental Analysis of Human Behavior Bulletin, 30*, 1-5.

Chapman, G. B. (1996). Temporal discounting and utility for health and money. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(3)*, 771-791. https://doi.org/10.1037/0278-7393.22.3.771

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization, 81(1)*, 1-8. https://doi.org/10.1016/j.jebo.2011.08.009

Cheng, Y. Y., Shein, P. P., & Chiou, W. B. (2012). Escaping the impulse to immediate gratification: The prospect concept promotes a future-oriented mindset, prompting an inclination towards delayed gratification. *British Journal of Psychology, 103(1)*, 129-141. https://doi.org/10.1111/j.2044-8295.2011.02067.x

*Chiou, W. B., & Wu, W. H. (2016). Episodic future thinking involving the nonsmoking self can induce lower discounting and cigarette consumption. *Journal of Studies on Alcohol and Drugs, 78(1)*, 106-112. https://doi.org/10.15288/jsad.2017.78.106

Church, R. M., & Deluty, M. Z. (1977). Bisection of temporal intervals. Journal of Experimental Psychology: *Animal Behavior Processes, 3*(3), 216-228. https://doi.org/10.1037//0097-7403.3.3.216

Cumming, G. (2011). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Dai, Z., Grace, R. C., & Kemp, S. (2009). Reward contrast in delay and probability discounting. *Learning & Behavior, 37(3)*, 281-288. https://doi.org/10.3758/LB.37.3.281

Dallery, J., & Raiff, B. R. (2007). Delay discounting predicts cigarette smoking in a laboratory model of abstinence reinforcement. *Psychopharmacology, 190*, 485-496. https://doi.org/10.1007/s00213-006-0627-5

*Daniel, T. O., Said, M., Stanton, C. M., & Epstein, L. H. (2015). Episodic future thinking reduces delay discounting and energy intake in children. *Eating Behaviors, 18*, 20-24. https://doi.org/10.1016/j.eatbeh.2015.03.006

*Daniel, T. O., Sawyer, A., Dong, Y., Bickel, W. K., & Epstein, L. H. (2016). Remembering versus imagining: When does episodic retrospection and episodic prospection aid decision making? *Journal of Applied Research in Memory and Cognition, 5(3),* 352-358. https://doi.org/10.1016/j.jarmac.2016.06.005

*Daniel, T. O., Stanton, C. M., & Epstein, L. H. (2013a). The future is now: comparing the effect of episodic future thinking on impulsivity in lean and obese individuals. *Appetite, 71*, 120-125. https://doi.org/10.1016/j.appet.2013.07.010

*Daniel, T. O., Stanton, C. M., & Epstein, L. H. (2013b). The future is now: reducing impulsivity and energy intake using episodic future thinking. *Psychological Science, 24(11)*, 2339-2342. https://doi.org/10.1177/0956797613488780

D'Argembeau, A., Renaud, O., & Van der Linden, M. (2011). Frequency, characteristics and functions of future-oriented thoughts in daily life. *Applied Cognitive Psychology, 25*(1), 96-103. https://doi.org/10.1002/acp.1647

D'Argembeau, A., & Van der Linden, M. (2004). Phenomenal characteristics associated with projecting oneself back into the past and forward into the future: Influence of valence and temporal distance. *Consciousness and Cognition, 13*(4), 844-858. https://doi.org/10.1016/j.concog.2004.07.007

*Dassen, F. C., Jansen, A., Nederkoorn, C., & Houben, K. (2016). Focus on the future: episodic future thinking reduces discount rate and snacking. *Appetite, 96,* 327-332. https://doi.org/10.1016/j.appet.2015.09.032

Davis, C., Patte, K., Curtis, C., & Reid, C. (2010). Immediate pleasures and future consequences. A neuropsychological study of binge eating and obesity. *Appetite*, *54(1)*, 208-213. https://doi.org/10.1016/j.appet.2009.11.002

De La Vega, A. (2011). *Does thinking about the past reduce temporal discounting? An investigation into the effects of episodic thought on intertemporal choice*. Unpublished master's dissertation, University of Colorado.

De Wit, H. (2009). Impulsivity as a determinant and consequence of drug use: a review of underlying processes. *Addiction Biology, 14*, 22-31. https://doi.org/10.1111/j.1369-1600.2008.00129.x

Dixon, M. R., Marley, J., & Jacobs, E. A. (2003). Delay discounting by pathological gamblers. *Journal of Applied Behavior Analysis, 36(4)*, 449-458. https://doi.org/10.1901/jaba.2003.36-449

Du, W., Green, L., & Myerson, J. (2002). Cross-cultural comparisons of discounting delayed and probabilistic rewards. *Psychological Record, 52(4)*, 479-492. https://doi.org/10.1007/BF03395199

Duval, S.J. and Tweedie, R.L. (1998). Practical estimates of the effect of publication bias in meta-analysis. *Australasian Epidemiologist, 5,* 14–17.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315(7109)*, 629-634. https://doi.org/10.1136/bmj.315.7109.629

Epstein, L. H., Richards, J. B., Saad, F. G., Paluch, R. A., Roemmich, J. N., & Lerman, C. (2003). Comparison between two measures of delay discounting in smokers. *Experimental and Clinical Psychopharmacology, 11*(2), 131-138. https://doi.org/10.1037/1064-1297.11.2.131

Fernie, G., Cole, J. C., Goudie, A. J., & Field, M. (2010). Risk-taking but not response inhibition or delay discounting predict alcohol consumption in social drinkers. *Drug and Alcohol Dependence, 112*(1-2), 54-61. https://doi.org/10.1016/j.drugalcdep.2010.05.011

Field, A. (2009). *Discovering statistics using SPSS - 3rd edition*. London: Sage Publications.

Franco-Watkins, A. M., Rickard, T. C., & Pashler, H. (2010). Taxing executive processes does not necessarily increase impulsive decision making. *Experimental Psychology, 57*, 193-201. https://doi.org/10.1027/1618-3169/a000024

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141(1)*, 2-18. https://doi.org/10.1037/a0024338

Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological Bulletin, 83*(2), 314-320. https://doi.org/10.1037//0033-2909.83.2.314

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

Green, L., Fry, A. F., & Myerson, J. (1994). Discounting of delayed rewards: A life-span comparison. *Psychological Science, 5(1),* 33-36. https://doi.org/10.1111/j.1467-9280.1994.tb00610.x

Green, L., & Myerson, J. (1996). Exponential versus hyperbolic discounting of delayed outcomes: Risk and waiting time. *American Zoologist, 36(4)*, 496-505. https://doi.org/10.1093/icb/36.4.496

Green, L., & Myerson, J. (2010). Experimental and correlational analyses of delay and probability discounting. In G. J. Madden & W. K. Bickel (Eds.), *Impulsivity: Theory, science, and neuroscience of discounting (pp. 67–92).* Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/12069-003

Green, L., Myerson, J., & McFadden, E. (1997). Rate of temporal discounting decreases with amount of reward. *Memory & Cognition, 25(5)*, 715-723. https://doi.org/10.3758/BF03211314

Green, L., Myerson, J., & Ostaszewski, P. (1999). Discounting of delayed rewards across the life span: age differences in individual discounting functions. *Behavioural Processes, 46(1),* 89-96. https://doi.org/10.1016/S0376-6357(99)00021-2

*Gutsin, B. D., Bree, K. A., Paluch, R. A., Stein, J. S., Bickel, W. K., & Epstein, L. H. (2018). *Episodic Future Thinking and Delay Discounting in Persons with Prediabetes*. Unpublished poster, University at Buffalo.

Hinvest, N. S., & Anderson, I. M. (2010). The effects of real versus hypothetical reward on delay and probability discounting. *The Quarterly Journal of Experimental Psychology, 63(6),* 1072-1084. https://doi.org/10.1080/17470210903276350

Hirsh, J. B., Guindon, A., Morisano, D., & Peterson, J. B. (2010). Positive mood effects on delay discounting. *Emotion, 10*(5), 717-721. https://doi.org/10.1037/a0019466

Holt, D. D., Green, L., & Myerson, J. (2003). Is discounting impulsive?: Evidence from temporal and probability discounting in gambling and non-gambling college students. *Behavioural Processes, 64*(3), 355-367. https://doi.org/10.1016/S0376-6357(03)00141-4

Holt, D. D., Green, L., & Myerson, J. (2012). Estimating the subjective value of future rewards: Comparison of adjusting-amount and adjusting-delay procedures. *Behavioural Processes, 90*(3), 302-310. https://doi.org/10.1016/j.beproc.2012.03.003

*Hu, X., Kleinschmidt, H., Martin, J. A., Han, Y., Thelen, M., Meiberth, D., Jessen, F., & Weber, B. (2017a). A reduction in delay discounting by using episodic future imagination and the association with episodic memory capacity. *Frontiers in Human Neuroscience, 10*, 663. https://doi.org/10.3389/fnhum.2016.00663

*Hu, X., Uhle, F., Fliessbach, K., Wagner, M., Han, Y., Weber, B., & Jessen, F. (2017b). Reduced future-oriented decision making in individuals with subjective cognitive decline: A functional MRI study. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 6*, 222-231. https://doi.org/10.1016/j.dadm.2017.02.005

Johnson, M. W., & Bickel, W. K. (2002). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the Experimental Analysis of Behavior, 77(2)*, 129-146. https://doi.org/10.1901/jeab.2002.77-129

Johnson, M. W., & Bickel, W. K. (2008). An algorithm for identifying nonsystematic delay-discounting data. *Experimental and Clinical Psychopharmacology*, *16(3)*, 264-274. https://doi.org/10.1037/1064-1297.16.3.264

Johnson M. W., Bickel W. K., & Baker F. (2007) Moderate drug use and delay discounting: A comparison of heavy, light, and never smokers. *Experimental and Clinical Psychopharmacology, 15*, 187–194. https://doi.org/10.1037/1064-1297.15.2.187

Johnson, P. S., Herrmann, E. S., & Johnson, M. W. (2015). Opportunity costs of reward delays and the discounting of hypothetical money and cigarettes. *Journal of the Experimental Analysis of Behavior*, *103*, 87–107. https://doi.org/10.1002/jeab.110

Joireman, J., Shaffer, M. J., Balliet, D., & Strathman, A. (2012). Promotion orientation explains why future-oriented people exercise and eat healthy: Evidence from the two-factor consideration of future consequences-14 scale. *Personality and Social Psychology Bulletin*, *38*(10), 1272-1287. https://doi.org/10.1177/0146167212449362

Jones, B., & Rachlin, H. (2006). Social discounting. *Psychological Science, 17*(4), 283-286. https://doi.org/10.1111/j.1467-9280.2006.01699.x

Jones, B. A., & Rachlin, H. (2009). Delay, probability, and social discounting in a public goods game. *Journal of the Experimental Analysis of Behavior, 91*(1), 61–73. https://doi.org/10.1901/jeab.2009.91-61

*Kakoschke, N., Hawker, C., Castine, B., de Courten, B., & Verdejo-Garcia, A. (2018). Smartphone-based cognitive bias modification training improves healthy food choice in obesity: A pilot study. *European Eating Disorders Review, 26(5)*, 526-532. https://doi.org/10.1002/erv.2622

Kim, H-Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry and Endodontics 38*, 52–54. https://doi.org/10.5395/rde.2013.38.1.52

*Kim, H., Schnall, S., & White, M. P. (2013). Similar psychological distance reduces temporal discounting. *Personality and Social Psychology Bulletin, 39(8)*, 1005-1016. https://doi.org/10.1177/0146167213488214

Kim, B. K., & Zauberman, G. (2013). Can Victoria's Secret change the future? A subjective time perception account of sexual-cue effects on impatience. *Journal of Experimental Psychology: General, 142*(2), 328-335. https://doi.org/10.1037/a0028954

Kirby, K. N., Winston, G. C., & Santiesteban, M. (2005). Impatience and grades: Delay-discount rates correlate negatively with college GPA. *Learning and Individual Differences*, *15(3)*, 213-222. https://doi.org/10.1016/j.lindif.2005.01.003

Klein, S. B., Loftus, J., & Kihlstrom, J. F. (2002). Memory and temporal experience: The effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. *Social Cognition, 20(5)*, 353-379. https://doi.org/10.1521/soco.20.5.353.21125

Koffarnus, M. N., & Bickel, W. K. (2014). A 5-trial adjusting delay discounting task: Accurate discount rates in less than one minute. *Experimental and Clinical Psychopharmacology, 22*(3), 222-228. https://doi.org/10.1037/a0035973

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15(2)*, 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kopec, C. D., & Brody, C. D. (2010). Human performance on the temporal bisection task. *Brain and Cognition, 74*(3), 262-272. https://doi.org/10.1016/j.bandc.2010.08.006

Krishnan-Sarin, S., Reynolds, B., Duhig, A. M., Smith, A., Liss, T., McFetridge, A., et al. (2007). Behavioral impulsivity predicts treatment outcome in a smoking cessation program for adolescent smokers. *Drug and Alcohol Dependence, 88*, 79-82. https://doi.org/10.1016/j.drugalcdep.2006.09.006

Kwan, D., Carson, N., Addis, D. R., & Rosenbaum, R. S. (2010). Deficits in past remembering extend to future imagining in a case of developmental amnesia. *Neuropsychologia, 48(11)*, 3179-3186. https://doi.org/10.1016/j.neuropsychologia.2010.06.011

Kwan, D., Craver, C.F., Green, L., Myerson, J., Boyer, P., Rosenbaum, R.S. (2012). Future decision-making without episodic mental time travel. *Hippocampus, 22,* 1215–1219. https://doi.org/10.1002/hipo.20981

Kwan, D., Craver, C. F., Green, L., Myerson, J., Gao, F., Black, S. E., & Rosenbaum, R. S. (2015). Cueing the personal future to reduce discounting in intertemporal choice: Is episodic prospection necessary? *Hippocampus, 25(4),* 432-443. https://doi.org/10.1002/hipo.22431

Lagorio, C. H., & Madden, G. J. (2005). Delay discounting of real and hypothetical rewards III: Steady-state assessments, forced-choice trials, and all real rewards. *Behavioural Processes, 69*(2), 173-187. https://doi.org/10.1016/j.beproc.2005.02.003

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4,* 863. https://doi.org/10.3389/fpsyg.2013.00863

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363-374. https://doi.org/10.2307/2529786

Laraway, S., Snycerski, S., Michael, J., & Poling, A. (2003). Motivating operations and terms to describe them: Some further refinements. *Journal of Applied Behavior Analysis, 36,* 407-414. https://doi.org/10.1901/jaba.2003.36-407

Lawrence, A. J., Luty, J., Bogdan, N. A., Sahakian, B. J., & Clark, L. (2009). Problem gamblers share deficits in impulsive decision-making with alcohol-dependent individuals. *Addiction, 104(6),* 1006-1015. https://doi.org/10.1111/j.1360-0443.2009.02533.x

*Lempert, K. M., Porcelli, A. J., Delgado, M. R., & Tricomi, E. (2012). Individual differences in delay discounting under acute stress: the role of trait perceived stress. *Frontiers in Psychology, 3,* 251. https://doi.org/10.3389/fpsyg.2012.00251

Lempert, K. M., Speer, M. E., Delgado, M. R., & Phelps, E. A. (2017). Positive autobiographical memory retrieval reduces temporal discounting. *Social Cognitive and Affective Neuroscience, 12(10),* 1584-1593. https://doi.org/10.1093/scan/nsx086

Li, X. (2007). The effects of appetitive stimuli on out-of-domain consumption impatience. *Journal of Consumer Research, 34*(5), 649-656. https://doi.org/10.1086/521900

*Lin, H., & Epstein, L. H. (2014). Living in the moment: Effects of time perspective and emotional valence of episodic thinking on delay discounting. *Behavioral Neuroscience, 128(1),* 12-19. https://doi.org/10.1037/a0035705

*Liu, L., Feng, T., Chen, J., & Li, H. (2013). The value of emotion: how does episodic prospection modulate delay discounting? *PLoS One, 8(11)*, e81717. https://doi.org/10.1371/journal.pone.0081717

Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *The Annals of the American Academy of Political and Social Science, 587,* 69-81. https://doi.org/10.1177/0002716202250791

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. California: Sage Publications, Inc.

Liviatan, I., Trope, Y., & Liberman, N. (2008). Interpersonal similarity as a social distance dimension: Implications for perception of others' actions. *Journal of Experimental Social Psychology, 44*(5), 1256-1269. https://doi.org/10.1016/j.jesp.2008.04.007

Madden, G. J., Begotka, A. M., Raiff, B. R., & Kastern, L. L. (2003). Delay discounting of real and hypothetical rewards. *Experimental and Clinical Psychopharmacology, 11(2),* 139-145. https://doi.org/10.1037/1064-1297.11.2.139

Madden, G. J., Bickel, W. K., & Jacobs, E. A. (1999). Discounting of delayed rewards in opioid-dependent outpatients: exponential or hyperbolic discounting functions? *Experimental and Clinical Psychopharmacology, 7(3),* 284-293. http://dx.doi.org/10.1037/1064-1297.7.3.284

Madden, G. J., & Johnson, P. S. (2010). A delay-discounting primer. In G. J. Madden, & W. K. Bickel (Eds.), *Impulsivity: The Behavioural and Neurological Science of Discounting* (pp. 11-37). Washington DC: American Psychological Association. https://doi.org/10.1037/12069-001

Madden, G. J., Petry, N. M., Badger, G. J., & Bickel, W. K. (1997). Impulsive and self-control choices in opioid-dependent patients and non-drug-using control participants: Drug and monetary rewards. *Experimental and Clinical Psychopharmacology, 5(3),* 256-262. https://doi.org/10.1037/1064-1297.5.3.256

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *The Effect of Delay and of*

Intervening Events on Reinforcement Value: Quantitative Analyses of Behavior (pp. 55–73). Psychology Press.

Mellis, A. M., Snider, S. E., Deshpande, H. U., LaConte, S. M., & Bickel, W. K. (2019). Practicing prospection promotes patience: repeated episodic future thinking cumulatively reduces delay discounting. *Drug and Alcohol Dependence, 204*, 107507. https://doi.org/10.1016/j.drugalcdep.2019.06.010

Mitchell, S. H. (1999). Measures of impulsivity in cigarette smokers and non-smokers. *Psychopharmacology, 146*, 455–464. https://doi.org/10.1007/PL00005491

Mitchell, J. M., Fields, H. L., D'Esposito, M., & Boettiger, C. A. (2005). Impulsive responding in alcoholics. *Alcoholism, Clinical and Experimental Research, 29*, 2158-2169. https://doi.org/10.1097/01.alc.0000191755.63639.4a

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *International Journal of Surgery, 8(5)*, 336-341. https://doi.org/10.1016/j.ijsu.2010.02.007

*Mok, J. N. Y. (2017). *Is It Time? Examining the Effects of Episodic Imagining on Reward Discounting*. Unpublished master's dissertation, York University.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7(1)*, 105-125. https://doi.org/10.1037/1082-989X.7.1.105

Myerson, J., & Green, L. (1995). Discounting of delayed rewards: Models of individual choice. *Journal of the Experimental Analysis of Behavior*, *64(3)*, 263-276. https://doi.org/10.1901/jeab.1995.64-263

Myerson, J., Green, L., Hanson, J. S., Holt, D. D., & Estle, S. J. (2003). Discounting delayed and probabilistic rewards: Processes and traits. *Journal of Economic Psychology, 24*(5), 619-635. https://doi.org/10.1016/S0167-4870(03)00005-9

Myerson, J., Green, L., & Warusawitharana, M. (2001). Area under the curve as a measure of discounting. *Journal of the Experimental Analysis of Behavior, 76(2)*, 235-243. https://doi.org/10.1901/jeab.2001.76-235.

O’Connell, G., Christakou, A., Haffey, A. T., & Chakrabarti, B. (2013). The Role of Empathy in Choosing Rewards from Another’s Perspective. *Frontiers in Human Neuroscience, 7*, 174. https://doi.org/10.3389/fnhum.2013.00174

*O'Donnell, S., Daniel, T. O., & Epstein, L. H. (2017). Does goal relevant episodic future thinking amplify the effect on delay discounting? *Consciousness and Cognition, 51*, 10-16. https://doi.org/10.1016/j.concog.2017.02.014

O’Donnell, S., Hollis-Hansen, K., & Epstein, L. (2019). Mix and match: An investigation into whether episodic future thinking cues need to match discounting delays in order to be effective. *Behavioral Sciences, 9*(1), 1. https://doi.org/10.3390/bs9010001

Odum A. L., & Baumann A. A. (2007) Cigarette smokers show steeper discounting of both food and cigarettes than money. *Drug and Alcohol Dependence, 91*, 293–296. https://doi.org/10.1016/j.drugalcdep.2007.07.004

Oettingen, G. (1996). Positive fantasy and motivation. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 236-259). New York: Guilford Press.

Oettingen, G. (1999). Free fantasies about the future and the emergence of developmental goals. In J. Brandtstadter & R. M. Lerner (Eds.), *Action and self-development: Theory and research through the lifespan.* Thousand Oaks, CA: Sage.

Oettingen, G., & Gollwitzer, P. M. (2009). Making goal pursuit effective: Expectancy-dependent goal setting and planned goal striving. In J. P. Forgas, R. Baumeister,, & D. M. Tice (Eds.), *The psychology of selfregulation: Cognitive, affective, and motivational processes* (pp. 127– 146). Philadelphia, PA: Psychology Press.

Oettingen, G., & Mayer, D. (2002). The motivating function of thinking about the future: Expectations versus fantasies. *Journal of Personality and Social Psychology, 83*(5), 1198. https://doi.org/10.1037/0022-3514.83.5.1198

Oettingen, G., Mayer, D., & Thorpe, J. (2010). Self-regulation of commitment to reduce cigarette consumption: Mental contrasting of future with reality. *Psychology and Health, 25*(8), 961-977. https://doi.org/10.1080/08870440903079448

Oettingen, G., & Wadden, T. A. (1991). Expectation, fantasy, and weight loss: Is the impact of positive thinking always positive? *Cognitive Therapy and Research, 15*(2), 167-175. https://doi.org/10.1007/BF01173206

Owens, T. E. (2015). *Effect of Mood Induction and Negative Urgency on Delay Discounting Performance in fMRI.* Unpublished doctoral dissertation, University of Kansas. (Doctoral dissertation, University of Kansas).

Olsen, R. (2016). *Delay Discounting of Reinforcer Value and Student Success*. Unpublished master's dissertation, Victoria University of Wellington.

Olsen, R. A., Macaskill, A. C., & Hunt, M. J. (2018). A measure of delay discounting within the academic domain. *Journal of Behavioral Decision Making, 31*, 522-534. https://doi.org/10.1002/bdm.2074

O'Neill, J., Daniel, T. O., & Epstein, L. H. (2016). Episodic future thinking reduces eating in a food court. *Eating Behaviors, 20,* 9-13. https://doi.org/10.1016/j.eatbeh.2015.10.002

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17(11),* 776-783. https://doi.org/10.1037/h0043424

Özbek, M., Bohn, A., & Berntsen, D. (2017). Imagining the personal past: Episodic counterfactuals compared to episodic memories and episodic future projections. Memory & cognition, 45(3), 375-389. https://doi.org/10.3758/s13421-016-0671-2

*Palombo, D. J., Keane, M. M., & Verfaellie, M. (2015). The medial temporal lobes are critical for reward-based decision making under conditions that promote episodic future thinking. *Hippocampus, 25(3),* 345-353. https://doi.org/10.1002/hipo.22376

Palombo, D. J., Keane, M. M., & Verfaellie, M. (2016). Using future thinking to reduce temporal discounting: Under what circumstances are the medial temporal lobes critical? *Neuropsychologia*, *89,* 437-444. https://doi.org/10.1016/j.neuropsychologia.2016.07.002

Parthasarathi, T., McConnell, M. H., Luery, J., & Kable, J. W. (2017). The vivid present: Visualization abilities are associated with steep discounting of future rewards. *Frontiers in Psychology*, *8*, 289. https://doi.org/10.3389/fpsyg.2017.00289

Perry, J. L., & Carroll, M. E. (2008). The role of impulsive behavior in drug abuse. *Psychopharmacology*, *200,* 1-26. https://doi.org/10.1007/s00213-008-1173-0

*Peters, J., & Büchel, C. (2010). Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions. *Neuron, 66(1)*, 138-148. https://doi.org/10.1016/j.neuron.2010.03.026

Peters, J., Miedl, S. F., & Büchel, C. (2012). Formal comparison of dual-parameter temporal discounting models in controls and pathological gamblers. *PloS one, 7(11)*, e47225. https://doi.org/10.1371/journal.pone.0047225

Petry, N. M. (2001). Delay discounting of money and alcohol in activelyusing alcoholics, currently abstinent alcoholics, and controls. *Psychopharmacology, 154*, 243–250. https://doi.org/10.1007/s002130000638

Petry, N. M., & Madden, G. J. (2010). Discounting and pathological gambling. In G. J. Madden and W. K. Bickel (Eds.), *Impulsivity: The behavioral and neurological science of discounting* (pp. 273–294). Washington, DC: American Psychological Association. https://doi.org/10.1037/12069-010

Pierce, W., & Cheney, C. (2017). *Behavior Analysis and Learning - Sixth Edition*. London: Routledge. https://doi.org/10.4324/9781315200682

Pronin, E., Olivola, C. Y., & Kennedy, K. A. (2008). Doing unto future selves as you would do unto others: Psychological distance and decision making. *Personality and Social Psychology Bulletin, 34*, 224-236. https://doi.org/10.1177/0146167207310023

Rachlin, H., & Jones, B. (2008). Social discounting and delay discounting. *Journal of Behavioural Decision Making, 21*, 29-43. https://doi.org/10.1002/bdm.567

Rachlin, H., & Jones, B. A. (2009). The extended self. In G. J. Madden & W. K. Bickel (Eds.), *Impulsivity: The behavioral and neurological science of discounting* (pp. 411–431). Washington DC: APA Books. https://doi.org/10.1037/12069-015

Rachlin, H., Locey, M. (2011). A behavioral analysis of altruism. *Behavioral Processes, 87*(1), 25–33. https:// doi.org/10.1016/j.beproc.2010.12.004

Rachlin, H., Raineri, A., & Cross, D. (1991). Subjective probability and delay. Journal of the *Experimental Analysis of Behavior, 55(2)*, 233-244. https://doi.org/10.1901/jeab.1991.55-233

Radu, P. T., Yi, R., Bickel, W. K., Gross, J. J., & McClure, S. M. (2011). A mechanism for reducing delay discounting by altering temporal attention. *Journal of the Experimental Analysis of Behavior, 96*(3), 363-385. https://doi.org/10.1901/jeab.2011.96-363

Read, D., Frederick, S., Orsel, B., & Rahman, J. (2005). Four score and seven years from now: The date/delay effect in temporal discounting. *Management Science, 51*(9), 1326-1335. https://doi.org/10.1287/mnsc.1050.0412

Reynolds, B. (2006). A review of delay-discounting research with humans: relations to drug use and gambling. *Behavioural Pharmacology, 17*, 651-667. https://doi.org/10.1097/FBP.0b013e3280115f99

Richards, F. J. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany*, *10*(2), 290–300. https://doi.org/10.1093/jxb/10.2.290

Rohatgi, A. (2015). WebPlotDigitizer (Version 4.1) [Computer software]. Retrieved from https://automeris.io/WebPlotDigitizer.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86(3),* 638-641. https://doi.org/10.1037/0033-2909.86.3.638

Rung, J. M., Argyle, T. M., Siri, J. L., & Madden, G. J. (2018). Choosing the right delay-discounting task: Completion times and rates of nonsystematic data. *Behavioural Processes, 151*, 119-125. https://doi.org/10.1016/j.beproc.2018.03.022

Rung, J. M., & Madden, G. J. (2018a). Demand characteristics in episodic future thinking: Delay discounting and healthy eating. *Experimental and Clinical Psychopharmacology*, *26(1)*, 77-84. https://doi.org/10.1037/pha0000171

Rung, J. M., & Madden, G. J. (2018b). Experimental reductions of delay discounting and impulsive choice: A systematic review and meta-analysis. *Journal of Experimental Psychology: General, 147(9),* 1349-1381. https://doi.org/10.1037/xge0000462

Rung, J. M., & Madden, G. J. (2019). Demand characteristics in episodic future thinking II: The role of cues and cue content in changing delay discounting. *Experimental and Clinical Psychopharmacology*, *27*(5), 482-495. http://dx.doi.org/10.1037/pha0000260

Sasse, L. K., Peters, J., & Brassen, S. (2017). Cognitive control modulates effects of episodic simulation on delay discounting in aging. *Frontiers in Aging Neuroscience, 9,* 58. https://doi.org/10.3389/fnagi.2017.00058

Sasse, L. K., Peters, J., Büchel, C., & Brassen, S. (2015). Effects of prospective thinking on intertemporal choice: the role of familiarity. *Human Brain Mapping, 36(10),* 4210-4221. https://doi.org/10.1002/hbm.22912

Sawyer, A., G. (1975). Detecting demand characteristics in laboratory experiments in consumer research: The case of repetition-affect research. *Advances in Consumer Research, 2*, 713–24.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B, 362*, 773–786. https://doi.org/10.1098/rstb.2007.2087

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience, 8*(9), 657-661. https://doi.org/10.1038/nrn2213

*Seidman, J., Hollis-Hansen, Aref, J., Hrynyk, I., & Epstein, L. *Improvements in Episodic Future Thinking Methodology: Identifying a Better Control*. Unpublished poster, University at Buffalo.

Sellitto, M., Ciaramelli, E., & di Pellegrino, G. (2010). Myopic discounting of future rewards after medial orbitofrontal damage in humans. *Journal of Neuroscience, 30*(49), 16429-16436. https://doi.org/10.1523/JNEUROSCI.2516-10.2010

Smith, C. L., & Hantula, D. A. (2008). Methodological considerations in the study of delay discounting in intertemporal choice: A comparison of tasks and modes. *Behavior Research Methods, 40(4),* 940-953. https://doi.org/10.3758/BRM.40.4.940

Smith, K. R., Lawyer, S. R., & Swift, J. K. (2018). A meta-analysis of nonsystematic responding in delay and probability reward discounting. *Experimental and Clinical Psychopharmacology, 26*(1), 94. https://doi.org/10.1037/pha0000167

*Snider, S. E., Deshpande, H. U., Lisinski, J. M., Koffarnus, M. N., LaConte, S. M., & Bickel, W. K. (2018). Working memory training improves alcohol users' episodic future thinking: A rate-dependent analysis. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3(2),* 160-167. https://doi.org/10.1016/j.bpsc.2017.11.002

*Snider, S. E., LaConte, S. M., & Bickel, W. K. (2016). Episodic future thinking: Expansion of the temporal window in individuals with alcohol dependence. *Alcoholism: Clinical and Experimental Research, 40(7)*, 1558-1566. https://doi.org/10.1111/acer.13112

Stein, J. S., Daniel, T. O., Epstein, L. H., & Bickel, W. K. (2015). Episodic future thinking reduces delay discounting in cigarette smokers. *Drug & Alcohol Dependence, 156*, e212. https://doi.org/10.1016/j.drugalcdep.2015.07.571

*Stein, J. S., Sze, Y. Y., Athamneh, L., Koffarnus, M. N., Epstein, L. H., & Bickel, W. K. (2017). Think fast: rapid assessment of the effects of episodic future thinking on delay discounting in overweight/obese participants. *Journal of Behavioral Medicine, 40(5),* 832-838. https://doi.org/10.1007/s10865-017-9857-8

*Stein, J. S., Tegge, A. N., Turner, J. K., & Bickel, W. K. (2018). Episodic future thinking reduces delay discounting and cigarette demand: an investigation of the good-subject effect. *Journal of Behavioral Medicine, 41(2),* 269-276. https://doi.org/10.1007/s10865-017-9908-1

*Stein, J. S., Wilson, A. G., Koffarnus, M. N., Daniel, T. O., Epstein, L. H., & Bickel, W. K. (2016). Unstuck in time: Episodic future thinking reduces delay discounting and cigarette smoking. *Psychopharmacology, 233(21-22),* 3771-3778. https://doi.org/10.1007/s00213-016-4410-y

Stone, C. (2018). *The Distance Effect: Species or Procedure?* Unpublished honours dissertation, Victoria University of Wellington.

Strathman, A., Gleicher, F., Boninger, D. S., & Edwards, C. S. (1994). The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology, 66*(4), 742-752. https://doi.org/10.1037/0022-3514.66.4.742

Sze, Y. Y., Daniel, T. O., Kilanowski, C. K., Collins, R. L., & Epstein, L. H. (2015). Web-based and mobile delivery of an episodic future thinking intervention for overweight and obese families: a feasibility study. *JMIR mHealth and uHealth, 3(4),* e97. https://doi.org/10.2196/mhealth.4603

*Sze, Y. Y., Stein, J. S., Bickel, W. K., Paluch, R. A., & Epstein, L. H. (2017). Bleak present, bright future: online episodic future thinking, scarcity, delay discounting, and food demand. Clinical *Psychological Science, 5(4),* 683-697. https://doi.org/10.1177/2167702617696511

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review, 110*(3), 403-421. https://doi.org/10.1037/0033-295X.110.3.403

Van den Bergh, B., Dewitte, S., & Warlop, L. (2008). Bikinis instigate generalized impatience in intertemporal choice. *Journal of Consumer Research, 35*(1), 85-97.

Verhulst, P. F. (1847). Deuxieme memoire sur la loi d'accroissement de la population. *Memoires de l'Academie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique, Series 2, 20,* 32.

Wallace, M. (1956). Future time perspective in schizophrenia. The Journal of Abnormal and *Social Psychology, 52*, 240–245. http://dx.doi.org/10.1037/h0039899

Washio, Y., Higgins, S. T., Heil, S. H., McKerchar, T. L., Badger, G. J., Skelly, J. M., & Dantona, R. L. (2011). Delay discounting is associated with treatment response among cocaine-dependent outpatients. *Experimental and Clinical Psychopharmacology, 19(3)*, 243-248. https://doi.org/10.1037/a0023617

Weatherly, J. N., & Derenne, A. (2011). Comparing delay discounting rates when using the fill-in-the-blank and multiple-choice methods. *The Journal of General Psychology, 138*(4), 300-318. https://doi.org/10.1080/00221309.2011.606442

Weller, R. E., Cook III, E. W., Avsar, K. B., & Cox, J. E. (2008). Obese women show greater delay discounting than healthy-weight women. *Appetite, 51(3),* 563-569. https://doi.org/10.1016/j.appet.2008.04.010

Wiehler, A., Bromberg, U., & Peters, J. (2015). The role of prospection in steep temporal reward discounting in gambling addiction. *Frontiers in Psychiatry*, *6*, 112. https://doi.org/10.3389/fpsyt.2015.00112

*Wiehler, A., Petzschner, F. H., Stephan, K. E., & Peters, J. (2017). Episodic tags enhance striatal valuation signals during temporal discounting in pathological gamblers. *Cognition and Behavior*, *4(3)*, e0159. https://doi.org/10.1523/ENEURO.0159-17.2017

Wilcox, R. (2012). *Introduction to robust estimation and hypothesis testing - 3rd edition*. Amsterdam: Academic Press. https://doi.org/10.1016/B978-0-12-386983-8.00010-X

Willeumier, K. C., Taylor, D. V., & Amen, D. G. (2011). Elevated BMI is associated with decreased blood flow in the prefrontal cortex using SPECT imaging in healthy adults. *Obesity, 19(5),* 1095-1097. https://doi.org/10.1038/oby.2011.16

Wilson, M., & Daly, M. (2004). Do pretty women inspire men to discount the future? *Biology Letters, 271,* S177–S179. https://doi.org/10.1098/rsbl.2003.0134

*Wu, W. H., Cheng, W., & Chiou, W. B. (2017). Episodic future thinking about the ideal self induces lower discounting, leading to a decreased tendency toward cheating. *Frontiers in Psychology, 8,* 287. https://doi.org/10.3389/fpsyg.2017.00287

Yi, R., Carter, A. E., Landes R. D. (2012) Restricted psychological horizon in active methamphetamine users: future, past, probability, and social discounting. *Behavioral Pharmacology, 23*(4), 358–66. https://doi.org/10.1097/FBP.0b013e3283564e11

Yi, R., Pickover, A., Stuppy-Sullivan, A. M., Baker, S., & Landes, R. D. (2016). Impact of episodic thinking on altruism. *Journal of Experimental Social Psychology, 65*, 74-81. https://doi.org/10.1016/j.jesp.2016.03.005

*Yi, R., Stuppy-Sullivan, A., Pickover, A., & Landes, R. D. (2017). Impact of construal level manipulations on delay discounting. *PloS one, 12*(15), e0177240. https://doi.org/10.1371/journal.pone.0177240

Yoon, J. H., Higgins, S. T., Heil, S. H., Sugarbaker, R. J., Thomas, C. S., & Badger, G. J. (2007). Delay discounting predicts postpartum relapse to cigarette smoking among pregnant women. *Experimental and Clinical Psychopharmacology, 15*(2), 176-186. https://doi.org/10.1037/1064-1297.15.2.186

*Zhang, S., Peng, J., Qin, L., Suo, T., & Feng, T. (2018). Prospective emotion enables episodic prospection to shift time preference. *British Journal of Psychology, 109*, 487–499. https://doi.org/10.1111/bjop.12284

Appendix A: Consideration of Future Consequences (CFC) Scale (Joireman, Shaffer, Balliet, & Strathman, 2012)

For each of the statements below, please indicate whether or not the statement is characteristic of you. If the statement is extremely uncharacteristic of you (not at all like you) please write a "1" to the left of the question; if the statement is extremely characteristic of you (very much like you) please write a "7" next to the question. And, of course, use the numbers in the middle if you fall between the extremes. Please keep the following scale in mind as you rate each of the statements below.

1. I consider how things might be in the future, and try to influence those things with my day to day behavior. (F)
2. Often I engage in a particular behavior in order to achieve outcomes that may not result for many years. (F)
3. I only act to satisfy immediate concerns, figuring the future will take care of itself. (I)
4. My behavior is only influenced by the immediate (i.e., a matter of days or weeks) outcomes of my actions. (I)
5. My convenience is a big factor in the decisions I make or the actions I take.(I)
6. I am willing to sacrifice my immediate happiness or wellbeing in order to achieve future outcomes. (F)
7. I think it is important to take warnings about negative outcomes seriously even if the negative outcome will not occur for many years. (F)
8. I think it is more important to perform a behavior with important distant consequences than a behavior with less important immediate consequences. (F)
9. I generally ignore warnings about possible future problems because I think the problems will be resolved before they reach crisis level. (I)
10. I think that sacrificing now is usually unnecessary since future outcomes can be dealt with at a later time. (I)
11. I only act to satisfy immediate concerns, figuring that I will take care of future problems that may occur at a later date. (I)
12. Since my day-to-day work has specific outcomes, it is more important to me than behavior that has distant outcomes. (I)
13. When I make a decision, I think about how it might affect me in the future. (F)

14. My behavior is generally influenced by future consequences. (F)

I = CFC-Immediate subscale item

F = CFC-Future subscale item

Appendix B: Worked example of the corrected-AUC calculation for the control condition from Experiment 1 based on data from one participant.

Data from one participant from Experiment 1

EFT data		
Delays (days)	Indifference point (as a proportion of the larger, delayed reward)	Hyperbolic *k*
5	0.95	0.007
11	0.95	
30	0.95	
88	0.45	
173	0.05	
385	0.85	
Control data		
Delays (days)	Indifference point (as a proportion of the larger, delayed reward)	Hyperbolic *k*
8	0.95	**0.028**
20	0.95	
59	0.15	
130	0.05	
279	0.05	
461	0.05	

Note: Bolded values are used in the worked example below

Worked example of one trapezoid calculation for the corrected-AUC for the control condition

Control indifference point at the first delay predicted by the hyperbola using the EFT delay:

$$V = \frac{\mathrm{A}}{1 + \mathrm{kD}} = \frac{1}{1 + 0.028 \times 5} = 0.88$$

Control indifference point at the second delay predicted by the hyperbola using the EFT delay:

$$V = \frac{\mathrm{A}}{1 + \mathrm{kD}} = \frac{1}{1 + 0.028 \times 11} = 0.77$$

Corrected trapezoid calculation for these delays for the control AUC:

$$Trapezoid\ area = (x_2 - x_1) \times \left(\frac{y_1 + y_2}{2}\right)$$

$$= (11 - 5) \times \left(\frac{0.88 + 0.77}{2}\right)$$

$$= 4.93$$

Appendix C: Meta-analysis coding manual

META-ANALYSIS CODING MANUAL

The effect of episodic future thinking on delay discounting rate

Rebecca A. Olsen

2018

Inclusion criteria:

1. Studies must compare the effects of episodic future thinking (EFT) vs. a control on delay discounting rate, using an experimental design (see definitions of EFT and control below).
2. Delay discounting rates need to be obtained from a task in which participants made choices between smaller, sooner rewards and larger, delayed rewards.
3. Unpublished data will also be included, obtained by searching for theses and emailing authors prevalent in the field.

Exclusion criteria

1. Studies that do not report sufficient quantitative data in order to estimate effect sizes (and authors cannot be contacted for this information).
2. Studies that do not meet the EFT/control definitions (see below).

Criteria for EFT

A manipulation where participants were required to imagine (or write about) personal future events. This manipulation must be episodic (a specific event) and have a future component. E.g., studies where participants only engaged in semantic future thinking or episodic past thinking were excluded.

Criteria for control

A control condition or group that is not an EFT manipulation that can be compared to the EFT manipulation. This could be imagining personal past events, recent events, semantic thinking, or 'no thinking' (i.e., no extra task) during the delay discounting task. See coding template for definitions of each control type.

Study Identification

Each study that meets the above criteria was given a *study number* (see provided list). If there are **several data-sets within a study** that do not contain overlapping samples (i.e., separate experiments with different participants), then the separate data-sets will be assigned their own number within the study, expressed as a decimal point (e.g., Experiment 1 of study 26 will be 26.1; Experiment 2 of study 26 will be 26.2).

SECTIONS OF THE CODING TEMPLATE:

Cover sheet:

Only one cover sheet will be completed for **each study**, regardless of the number of data-sets or experiments. There is also space on the cover sheet to write down any notes about the study (e.g., whether extra data still needs to be obtained from authors or whether multiple effect sizes coded contain overlapping samples (see below)).

General statistics/study descriptors form:

One form will be completed for **each data-set that does not contain overlapping samples.** For example, if there are multiple experiments within a study (i.e., separate experiments with different participants), then one form will be used for each experiment. General statistics/study descriptors forms include:

1. Study identification number
2. Type of document
3. Whether additional information was obtained to code effect sizes (see provided list of study numbers). This is to ensure that the coder was aware of and referred to any additional information when coding the study.
4. The year the study was published/completed.
5. The country where the study originated (for currency conversion purposes)
6. The sample size
7. Average age of participants

   If several ages are reported within the same data-set then calculate the weighted average by multiplying each age by its N, adding these together then dividing by the total N:
   E.g., "control group average age = 28; control group N = 56; EFT group average age = 34; EFT group N = 42."

   $$weighted\ average = \frac{((28\times56)+(34\times42))}{(56+42)} = 30.57$$

8. Type of participants (e.g., are they healthy/a general population, addicts, amnesiacs, or obese?)
9. Type of reward (real, potentially real, or hypothetical – see definitions on coding template)
10. The type of discounting rate measure that was used
11. Whether a within-subjects or between-subjects design was used
12. The type of control group/manipulation used.

13. The delayed amount that was used
14. The largest delay that was used

<u>Effect size coding form</u>

While only one general statistics/study descriptors form is to be completed per independent data-set, a separate effect size coding form will be completed for each variable that we can code an effect size from.

<u>Multiple effect sizes where observations are not independent (i.e., overlapping samples):</u>

If multiple effect sizes are able to be coded within a study where observations are not independent (e.g., where there is multiple control groups or measures), a decision will be made later regarding which effect size to use in the meta-analysis (if these can't be averaged). However, all information will be coded initially. For example, if both *k* and AUC measures were used then an effect size will be coded for each measure, using two separate effect size coding forms (copy/paste a second effect size form within the document). The coder will also need to make a note on the cover sheet that these effect sizes were not statistically-independent.

Another example: If the same group of participants completed an EFT condition and two different control conditions (e.g., semantic and past thinking), then there are two possible effect sizes that should be coded (EFT/semantic and EFT/past) using two separate effect size coding forms. In both examples, the two effect sizes contain overlapping samples; code both effect sizes but make a note of this.

<u>Independent and dependent variables:</u>

Included at the top of each effect size coding form are the independent variable (e.g., EFT vs. no thinking) and the dependent variable or measure. The **independent variable** will be the specified EFT vs control conditions (see definitions above). Also include enough information here so that it is clear which effect size from the study the form corresponds to. The **dependent variable** will be the delay discounting rate measure (e.g., *k* value, AUC or the percent of instances in which the smaller, sooner option was chosen).

Coders will also code the page number and table (if applicable) from which the information came. This information will assist us when comparing effect sizes, to ensure we are coding from the same information, or alternatively to have a discussion about which data best fits our definition of the variable (if both coders used different measures to calculate the effect size).

Calculate as many effect sizes from a study as you think is appropriate:

These are some **examples** of the different variables that we can code an effect size from:

1. VALENCE
    1. Positive episodic future thinking
    2. Neutral episodic future thinking
    3. Negative episodic future thinking
2. IMAGERY TYPE e.g.,
    1. General/unrestricted
    2. Food-related
    3. Financial-goal-related
3. PARTICIPANT TYPE, e.g.,
    1. Amnesiacs
    2. Healthy/general population
    3. Obese/overweight
4. AMOUNT e.g.,
    1. $100
    2. $1,000

For example, if there are three effect sizes from a study that correspond to different valences, then specify under 'independent variable' which valence you have coded an effect size for on each coding form (positive, negative, or neutral episodic future thinking).

For example, if the amount of the larger delayed reward was manipulated (e.g., $100 vs. $1,000), then specify under 'independent variable' which amount you have coded an effect size for on each coding form ($100 or $1,000).

Effect size calculations (also see Appendix and excel calculator):

Effect sizes (Cohen's *d*) of the difference in discounting rate between EFT and control conditions. The final calculated effect sizes should be rounded to 3dp.

Ms, SDs, and Ns for each group or condition should be used to calculate effect sizes when available. Otherwise, effect sizes should be calculated from t test (or F test) statistics (see Appendix and excel calculator).

For **within-subjects designs**, Cohen's d_W will be converted to a between-subjects metric (d_{WC}) using the equation suggested by Morris and Deshon (2002) – see appendix and excel calculator. The correlation between EFT and control conditions (*r*) is always needed for this transformation. When *r* was unavailable, the author was contacted. If the author did not respond (see list of study numbers for this information), then *r* can be estimated from means, SDs and t tests when available (see appendix and excel calculator).

If an author reported their own Cohen's *d*, then the coders will still calculate this themselves based on the available information to ensure that the right calculation is used. This is particularly important for within-subjects designs as authors often do not specify their equations used so we do not know if the correlation has been taken into account. There are also other possible versions of Cohen's *d* for between-subjects designs that would not be appropriate to use for the current meta-analysis.

Negative effect sizes indicate that the EFT condition had a lower value than the control condition (this is consistent with the EFT effect if *k* value measures were used). Positive effect sizes indicate that the EFT condition had a higher value than the control condition (this is consistent with the EFT effect if AUC measures were used). **Where the effect size is consistent with the EFT effect (i.e., more self-control/shallower discounting in the EFT condition), final effect sizes should be reported as positive.**

Coders will also calculate the sampling **variance** of each effect size (see Appendix and excel calculator).

Below is the coding template (made up of the cover sheet, the general statistics/study descriptors form, and the effect size coding form) - one will be created for each listed study.

COVER SHEET EFTDD (2018)

STUDY NUMBER: ____

TITLE: ____

____

AUTHOR(S):____

____

____

____

YEAR(S):____

SOURCE (E.G., JOURNAL TITLE): ___________________________

____

____

NOTES (e.g., is additional info still necessary? Were observations NOT independent?):

____

____

____

____

____

VARIABLES EXCLUDED (e.g., SOME OTHER MEASURE OF NO INTEREST):

____

____

No. of effect sizes coded:____________________________

Date: ______________

Coder: _____________

GENERAL STATISTICS/STUDY DESCRIPTORS (One for each dataset with non-overlapping samples)

* STUDY - Identification #: ____

* DOCUMENT – What type of document? ___

0 = published (published in a journal)

1 = unpublished (not published in a journal article e.g., theses, manuscripts currently under review for publication, conference posters)

* ADDINFO – was additional information obtained to code effect sizes? ______

0 = no

1 = yes

* YEAR - date which study was published/released/completed: _____

* COUNTRY - country where study originated (for money conversions): ___

0 = Canada	9 = Scotland
1 = U.S.	10 = Sweden
2 = United Kingdom	11 = Austria
3 = France	12 = New Zealand
4 = Netherlands (Holland)	13 = Belgium
5 = Germany	14 = Taiwan
6 = Denmark	15 = China
7 = Norway	16 = Switzerland
8 = Australia	

* SIZE – sample size (analyzed):______

(Report overall sample size. The N according to groups etc. is taken into account on the effect size coding form)

* AVERAGE AGE – (years; 2dp): ______

(If EFT/control groups are reported separately then calculate the weighted average. Report for each independent effect size.)

* PPTTYPE – type of participants: ____

0 = healthy/general population

1 = alcoholics (including recovering)

2 = nicotine dependent (including recovering)

3 = other drug dependence (including recovering)

4 = amnesiacs

5 = obese/overweight

6 = other (specify): ______

* REWARD – Were the rewards real, potentially real or hypothetical? _____

0 = real (all rewards and delays experienced)

1 = potentially real (only one or some rewards and delays experienced, e.g., one outcome randomly selected and given to participant for real)

2 = hypothetical (no rewards or delays experienced)

* DRM – Discounting rate measure (AUC or *k* or other measure used?): _______

0 = AUC

1 = *k*

3 = other (specify): ____________________________

* Within subjects or between subjects design: _______

i.e., was EFT vs. control manipulated within or between?

0 = between

1 = within

* CONTROL – no thinking, past thinking, present thinking, etc.: _____

0 = no thinking (i.e., no extra task during the delay discounting task or a prior task

that does not target EFT components)

1 = past (exceeds 24 hours)

2 = present/recent/within 24 hours

3 = semantic/non-episodic (generalized information or an activity not localized in time)

4 = events from a story

5 = other (specify):____

* DAMOUNT – Delayed amount with currency (or specify range if various): ______

The amount of the larger later e.g., $100 now or $200 in x, delayed amount = $200.

* LDELAY – Largest delay used: ________________

EFFECT SIZE CODING – EFTDD – 2018 (copy/paste this form for each effect size)

Independent variable (e.g., positive EFT vs. no thinking): ________ Study No.:_______

Dependent variable (e.g., AUC): ____

Page #: ________________ Table: ______________

	Sample size	Mean	Standard deviation	Standard error (if no standard deviation available)
EFT				
Control				
t =	Total n =	r =		

NOTES: Specify in brackets if any numbers were estimated (such as t or r)

Effect size *d* (3dp): _______Statistic (2 is preferred if available): _____ Variance (3dp):________

***include negative sign for effect size if less self-control (i.e., "higher discount rate" or higher *k* value or lower AUC or chose smaller sooner more) in EFT condition**

Have you considered whether a negative effect size would be appropriate (yes/no)? _____

Possible statistics:

1. raw data 2. means (SD) 3. Estimated means and SDs from figure

4. t/F/p 5. *d* 6. Z (wilcoxon; change t to Z in table)

Notes on control type:

Coding manual Appendix

Equations for Calculating/Transforming Effect Sizes and Variance

*Note: all equations are in the excel spreadsheet calculator.

Cohen's *d* from sample sizes (*N*), means (*M*), and standard deviations (*SD*). Only appropriate for normally distributed data. Always try to calculate *d* this way first.

Between subjects (d_B):

$$d_B = \frac{M_1 - M_2}{S_W}$$

S_W is the pooled within standard deviation. $S_w = \sqrt{\frac{(N_1 - 1)(SD_1)^2 + (N_2 - 1)(SD_2)^2}{N_1 - 1 + N_2 - 1}}$

Source: Lakens (2013); Morris and deShon (2002)

Between subjects pre-test/post-test design (d_{BPP}):

$$d_{BPP} = \frac{M_{D,E}}{\mathrm{SD}_{PRE,E}} - \frac{M_{D,C}}{\mathrm{SD}_{PRE,C}}$$

For this type of design, we calculate an effect size for each group then subtract the effect size for the control group from the effect size for the EFT group, as shown in the equation above.

$M_{D,E}$ = the difference between pre- and post-test means for the EFT group (M_{POST} minus M_{PRE})

$SD_{PRE,E}$ = the pre-test standard deviation for the EFT group

$M_{D,C}$ = the difference between pre- and post-test means for the control group (M_{POST} minus M_{PRE})

$SD_{PRE,E}$ = the pre-test standard deviation for the control group

Source: Morris and deShon (2002)

Within subjects (d_W; also requires *r*):

$$d_W = \frac{M_1 - M_2}{\sqrt{SD_1^2 + SD_2^2 - 2 \times r \times SD_1 \times SD_2}}$$

$d_{WC} = d_W \times \sqrt{2(1-r)}$ d_{WC} is d_W converted to a between-subjects metric, where r is the correlation coefficient between the EFT and control conditions.

Source: Lakens (2013); Morris and deShon (2002)

From *SE* to *SD*

SE = $\frac{SD}{\sqrt{n}}$ SD = $SE * \sqrt{n}$

Cohen's *d* from *F/t*-tests*:*

Between subjects:

$$d_B = t\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$$

Source: Lakens (2013); Morris and deShon (2002)

Note: ***t* value can be determined from the exact *p* value** (between subjects only):

If an author reports the exact p value from a t test but not the t value, then the t value can be determined from the p value using the inverse distribution function in excel (see excel spread sheet).

Source: Lipsey and Wilson (2001)

Within-subjects:

$$d_W = \frac{t}{\sqrt{N}}$$

$d_{WC} = d_W \times \sqrt{2(1-r)}$

Source: Lakens (2013); Morris and deShon (2002)

Converting *F* values to *t* values (between and within subjects):

$$t = \sqrt{F}$$

Note: This conversion is only appropriate when *df* = 1. That is, the ANOVA compares two means such as the reported *F* for a main effect of condition (EFT vs. control).

e.g., F (1, 60) = appropriate; F (4,60) = not appropriate.

This conversion is not appropriate for an ANCOVA.

Non-normal data: Cohen's *d* from effect size *r* (r_{ES})

1. *Calculate effect size r from a Wilcoxon or Mann Whitney U:*

$$r_{ES} = \frac{Z}{\sqrt{\mathrm{N}}}$$

Note that for a Wilcoxon (i.e., within subjects), N is the number of observations, not the number of cases. So if there were 30 participants, N = 30 x 2 = 60.

2. *Convert the effect size (r) to the Cohen's d metric:*

$$d = \frac{2r_{ES}}{\sqrt{1-r_{ES}^2}}$$

Source: Fritz, Morris and Richler (2012)

3. *If within subjects (i.e., a Wilcoxon was used), then convert Cohen's d to the between-subjects metric (as in all other instances):*

$d_{WC} = d_W \sqrt{2(1-r(s))}$, where $r(s)$ is the Spearman's rank (non-parametric) correlation.

Source: Lakens (2013); Morris and deShon (2002)

Estimating *r* (correlation between conditions; last resort if not reported/author cannot be contacted; requires *N*, *M*s, *SD*s, and *t* value):

$$r = \frac{SD_{pre}^2 + SD_{post}^2 - SD_D^2}{(2)(SD_{pre})(SD_{post})}$$

$$SD_D^2 = \frac{n(M_{post} - M_{pre})^2}{t^2}$$

Or: $$r = 1 - \frac{SD_D^2}{2SD_P^2}$$

Where SD_D^2 is the variance of difference scores and SD_P^2 is the pooled variance (see excel spreadsheet)

Source: Morris and deShon (2002)

<u>Variance equations:</u>

Between-subjects:

$$Var_{d_B} = \left(\left(\frac{1}{\left(\frac{(n1 \times n2)}{(n1 + n2)}\right)}\right) \times \left(\frac{((n1 + n2) - 2)}{((n1 + n2) - 4)}\right) \times \left(1 + \left(\frac{(n1 \times n2)}{(n1 + n2)}\right) \times (d_B^2)\right)\right) - \left(\frac{(d_B^2)}{\left(\left(1 - \frac{3}{(4 \times ((n1 + n2) - 2) - 1)}\right)^2\right)}\right)$$

Source: Morris and deShon (2002)

Between-subjects pre-test/post-test design:

1. Calculate variance **for each group** using this equation:

$$Var_{d_{BPP}} = \left(\left(\frac{2\ (1 - r)}{n}\right) \times \left(\frac{(n - 1)}{(n - 3)}\right) \times \left(1 + \left(\frac{n}{(2(1 - r))}\right)\ (d^2)\right)\right) - \left(\frac{(d^2)}{\left(\left(1 - \frac{3}{(4 \times (n - 1) - 1)}\right)^2\right)}\right)$$

2. Sum the two variances together.

Source: Morris and deShon (2002)

Within-subjects:

$$Var_{d_{WC}} = \left(\left(\frac{1}{n}\right) \times \left(\frac{(n - 1)}{(n - 3)}\right) \times (1 + n \times (d_{WC}^2))\right) - \left(\frac{(d_{WC}^2)}{\left(\left(1 - \frac{3}{(4 \times (n - 1) - 1)}\right)^2\right)}\right)$$

Source: Morris and deShon (2002)

Coding manual references

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2-18.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. California: Sage Publications, Inc.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7(1), 105-125.

Appendix D: Description of initial coding by me and the first research assistant for the meta-analysis

Here we report the initial inter-rater reliability (IRR) between me and the first research assistant for the first 27 studies, given that independent re-checks and refinements to the coding manual were warranted prior to the training of the second research assistant (as described in the Method section of the meta-analysis).

I developed, tested and refined the coding manual and trained the research assistant on its use. Five studies were used for training purposes and IRR analyses (Absolute Agreement Intra-class correlations [ICCs] using a Two-Way Mixed model for continuous variables and Cohen's Kappa for categorical variables) is reported for the remaining 22.

We coded 38 common effect sizes, with good levels of agreement (ICC based on a single rater = 0.842; Koo & Li, 2016). Thirteen additional effect sizes were identified and coded by one rater but not the other. For the other continuous variables coded (n = 4), ICC values ranged from 0.120 to 1 (median ICC = 0.999). Effect size variance had a poor ICC value of 0.120; without this variable ICC ranged from 0.999 to 1, indicating excellent IRR. For categorical variables (n = 7), Cohen's Kappa ranged from 0.251 to 1 (median ICC = 1). Reward type had the lowest Cohen's Kappa of 0.251; without this variable Cohen's Kappa ranged from 0.789 to 1 (substantial to perfect; Landis & Koch, 1977).

We developed a clearer definition of reward type category and then independently rechecked coding of effect size variance and reward type variables. Following the independent re-checks, IRR was excellent for both effect size variance (ICC = 0.971) and reward type (Kappa = 1). IRR of effect sizes also increased after each rater independently re-checked whether any additional effect sizes could be coded (0.842 vs. 0.977). Following the re-checks there were 43 common effect sizes and 9 (compared to the initial 13) additional effect sizes that were identified and coded by one rater but not the other. Effect size and variance calculation disagreements were mostly due to rounding errors, and a consensus rating was reached for any disagreements for all variables. Both raters were unable to code an effect size for three studies due to insufficient quantitative data (Kwan et al., 2015; Sasse et al., 2015; Sasse et al., 2017).

As described in the Method section of the meta-analysis, I also coded the remaining 11 studies. However, the second rater for these studies was a second research assistant. The second research assistant was trained using the same studies as the first research assistant, and using the refined version of the standardized coding manual (i.e., containing the clearer definition of reward type developed by the first research assistant and me). Comparisons between my ratings and the two research assistants' ratings are reported in the Method section of the meta-analysis.

Appendix E: Plotted residuals for the hyperbolic model (Figure E.1) and logistic model (Figure E.2) for each condition from Experiments 2A, 2B and 2C

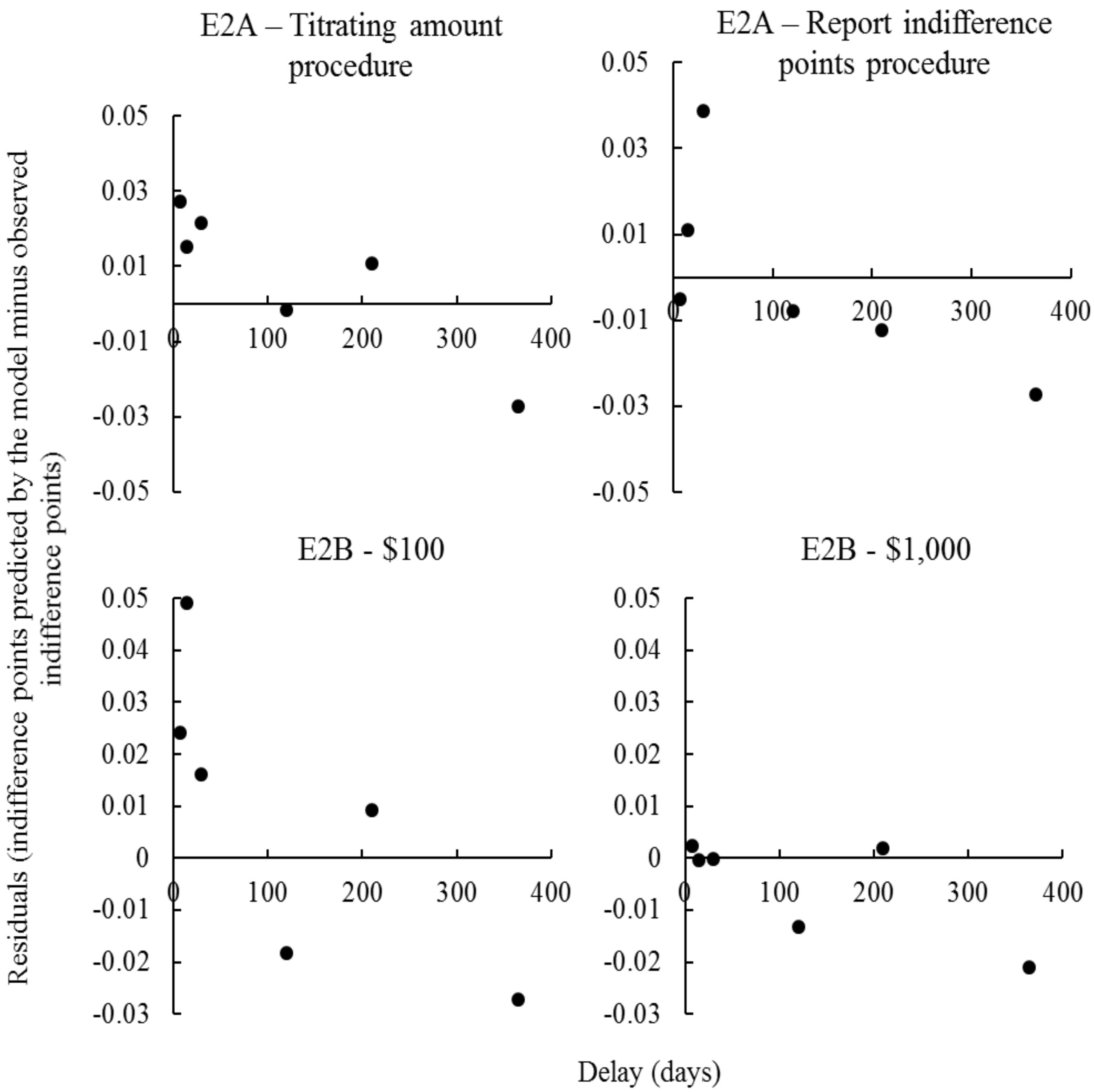

Figure E.1. Group median residuals based on systematic participants for the hyperbolic model for each delay in each condition from Experiments 2A and 2B. Points under the x axes represent underestimation of value by the hyperbolic model whereas points above the x axes represent overestimation of value by the hyperbolic model.

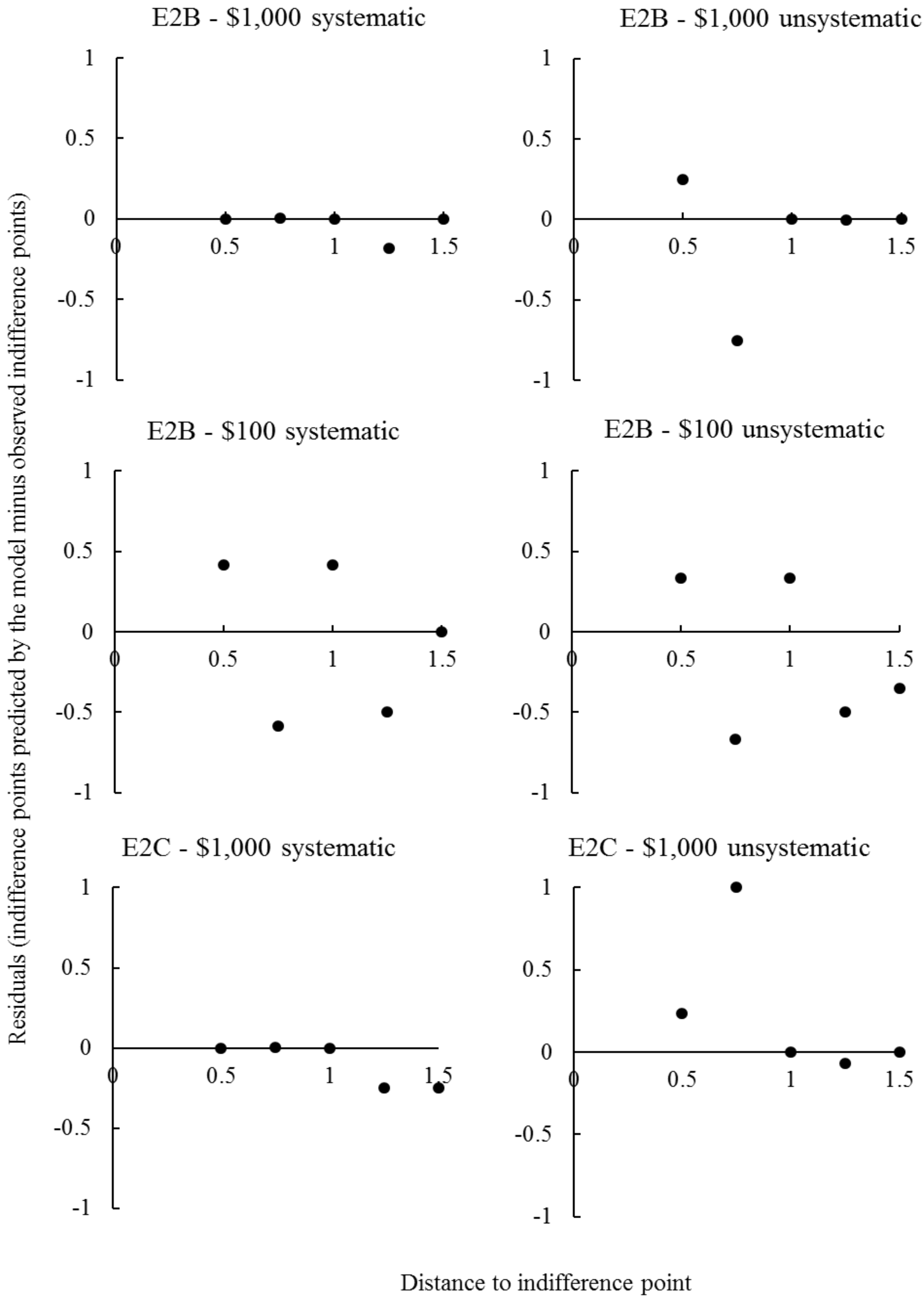

Figure E.2. Median residuals for the logistic model for each delay in each condition from Experiments 2A and 2B. Points under the x axes represent underestimation of value by the logistic model whereas points above the x axes represent overestimation of value by the logistic model.

Appendix F: Robust ANOVA results based on hyperbolic *k* values from Experiment 2A

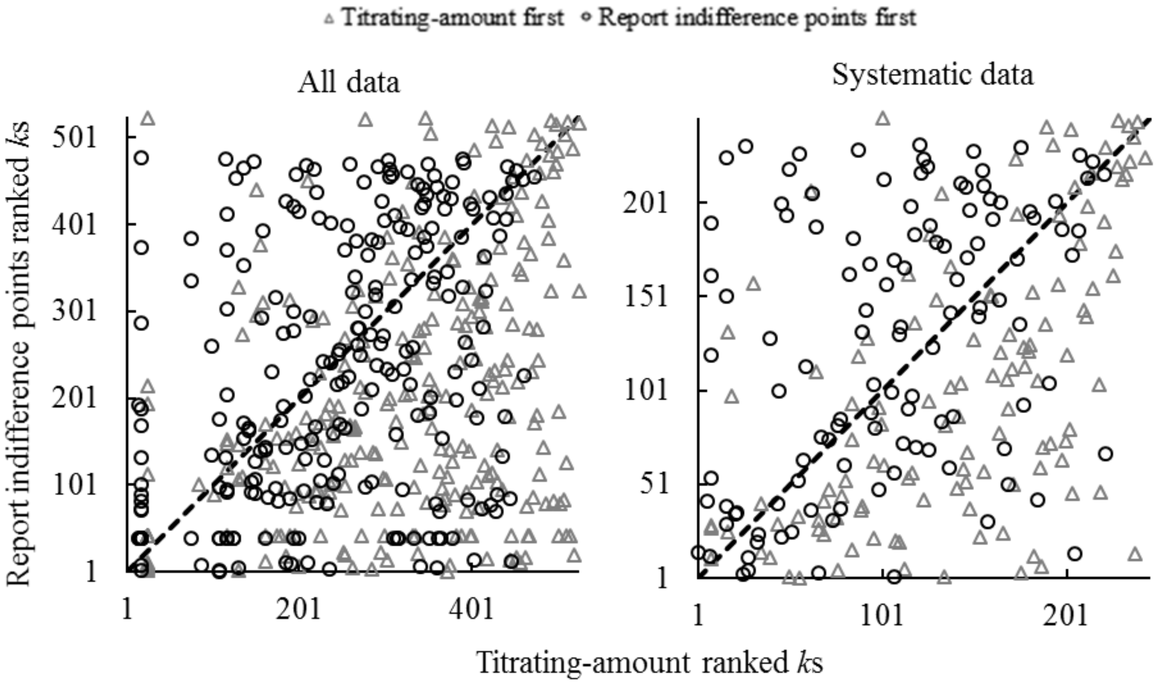

Figure F. Ranked *k* values for the titrating-amount procedure plotted against ranked *k* values for the report indifference points procedure. The ranks of the *k* values are plotted so that all values are visible despite the skewed distribution. Each plotted circle/triangle represents one individual; the black circles represent individuals who experienced the report indifference points procedure first, whereas the grey triangles represent individuals who experienced the titrating-amount procedure first. Plotted circles/triangles above the dashed line represent participants with higher *k*s in the report indifference points procedure, whereas plotted circles/triangles below the dashed line represent participants with higher *k*s in the titrating-amount procedure. The left graph is based on all data, whereas the right graph is only based on data from participants who showed systematic discounting.

The robust two-way mixed ANOVA based on all data (see left graph of Figure F) showed a significant main effect of order ($p = 0.023$) but no significant main effect of procedure ($p = 0.087$). There was a significant interaction between procedure and order ($p < 0.001$). Post hoc tests revealed that *k* values were significantly lower in the report indifference points procedure when the titrating-amount procedure was experienced first (report indifference points $M = 0.002$; titration $M = 0.008$; $p < 0.001$). However, there was no significant difference in *k* values when the report indifference points procedure was experienced first (report indifference points $M = 0.010$; titration $M = 0.007$; $p = 0.312$).

Different results were found based on systematic participants' data only (see right graph of Figure F). The robust two-way mixed ANOVA showed significant main effects of order ($p = 0.001$) and procedure ($p = 0.038$) and a significant interaction between procedure and order ($p < 0.001$). Although k values were significantly lower in the report indifference points procedure when the titrating-amount procedure was experienced first (report indifference points $M = 0.004$; titration $M = 0.009$; $p < 0.001$), k values were significantly *higher* in the report indifference points procedure when the report indifference points procedure was experienced first (report indifference points $M = 0.025$; titration $M = 0.010$; $p = 0.005$).

Appendix G: Figures depicting the logistic function fitted to data for unsystematic participants with R^2s < 0.5 from Experiments 2B (Figures G.1 and G.2) and 2C (Figure G.3).

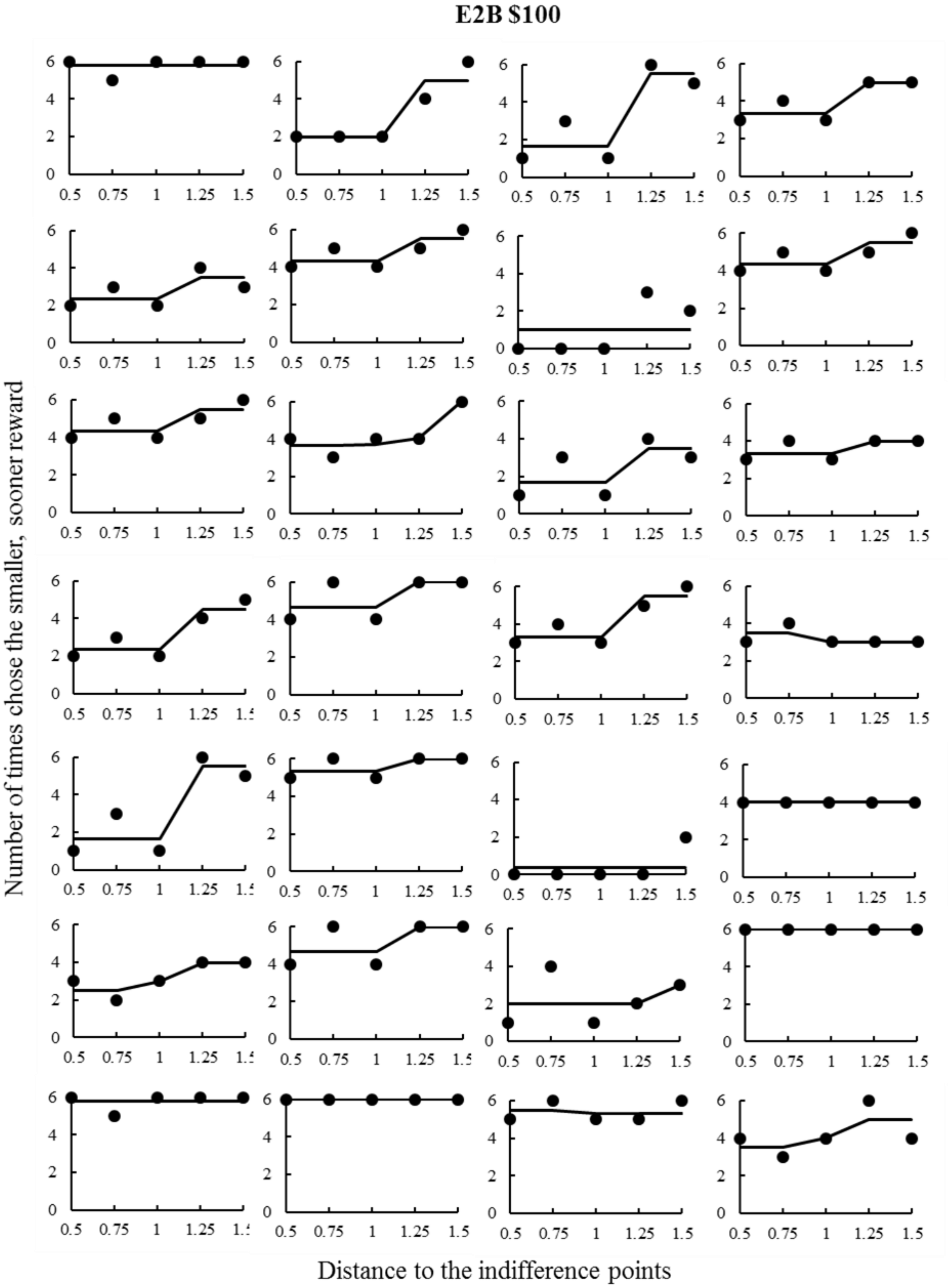

Figure G.1. The logistic function (black line) fitted to the number of times the smaller, sooner reward was chosen at each distance to the indifference point (black points) from Experiment 2B. Individuals who showed unsystematic discounting in the report indifference points procedure and had logistic model R^2s < 0.5 in the $100 condition in the distance to indifference points procedure are shown.

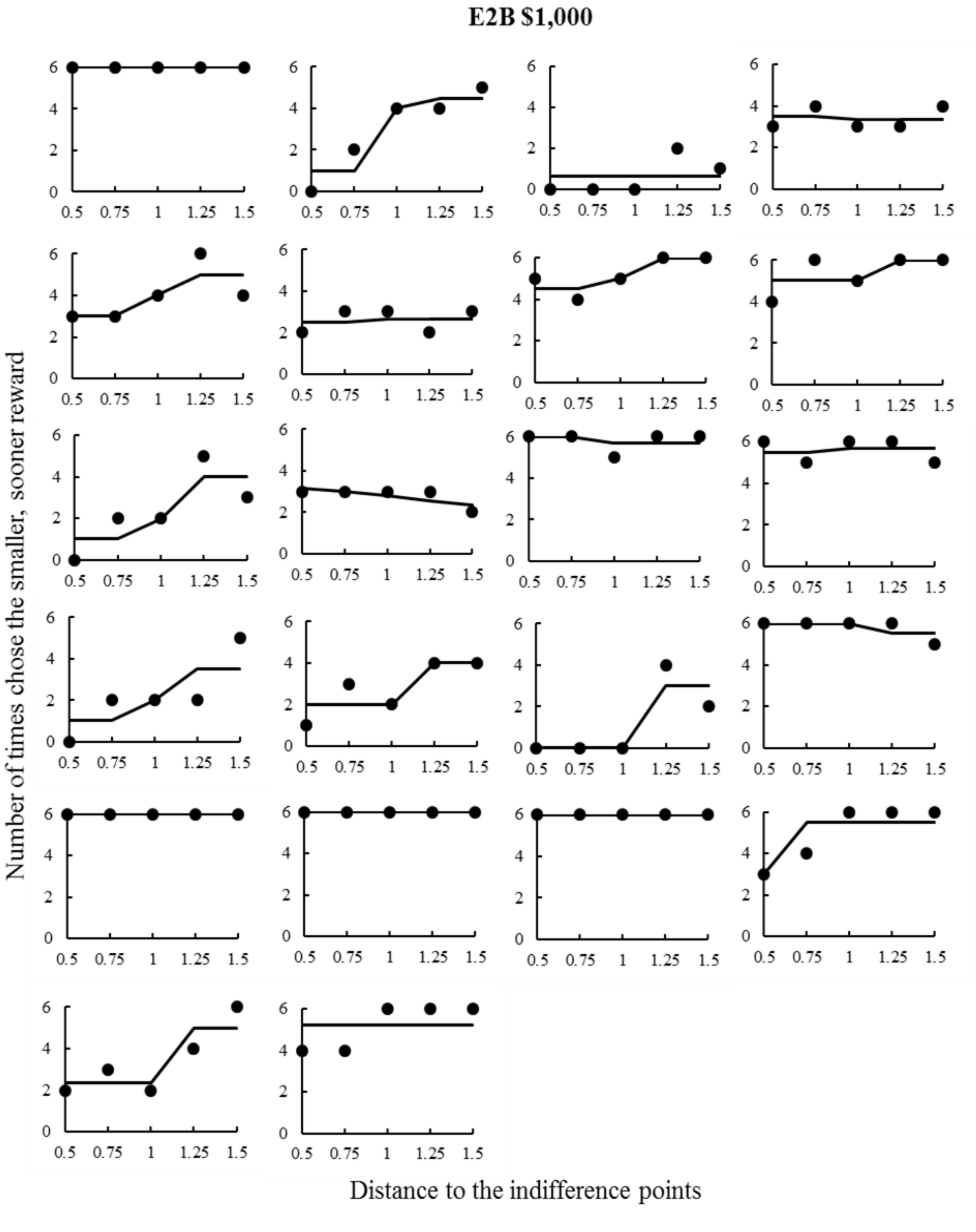

Figure G.2. The logistic function (black line) fitted to the number of times the smaller, sooner reward was chosen at each distance to the indifference point (black points) from Experiment 2B. Individuals who showed unsystematic discounting in the report indifference points procedure and had logistic model R^2s < 0.5 in the $1,000 condition in the distance to indifference points procedure are shown.

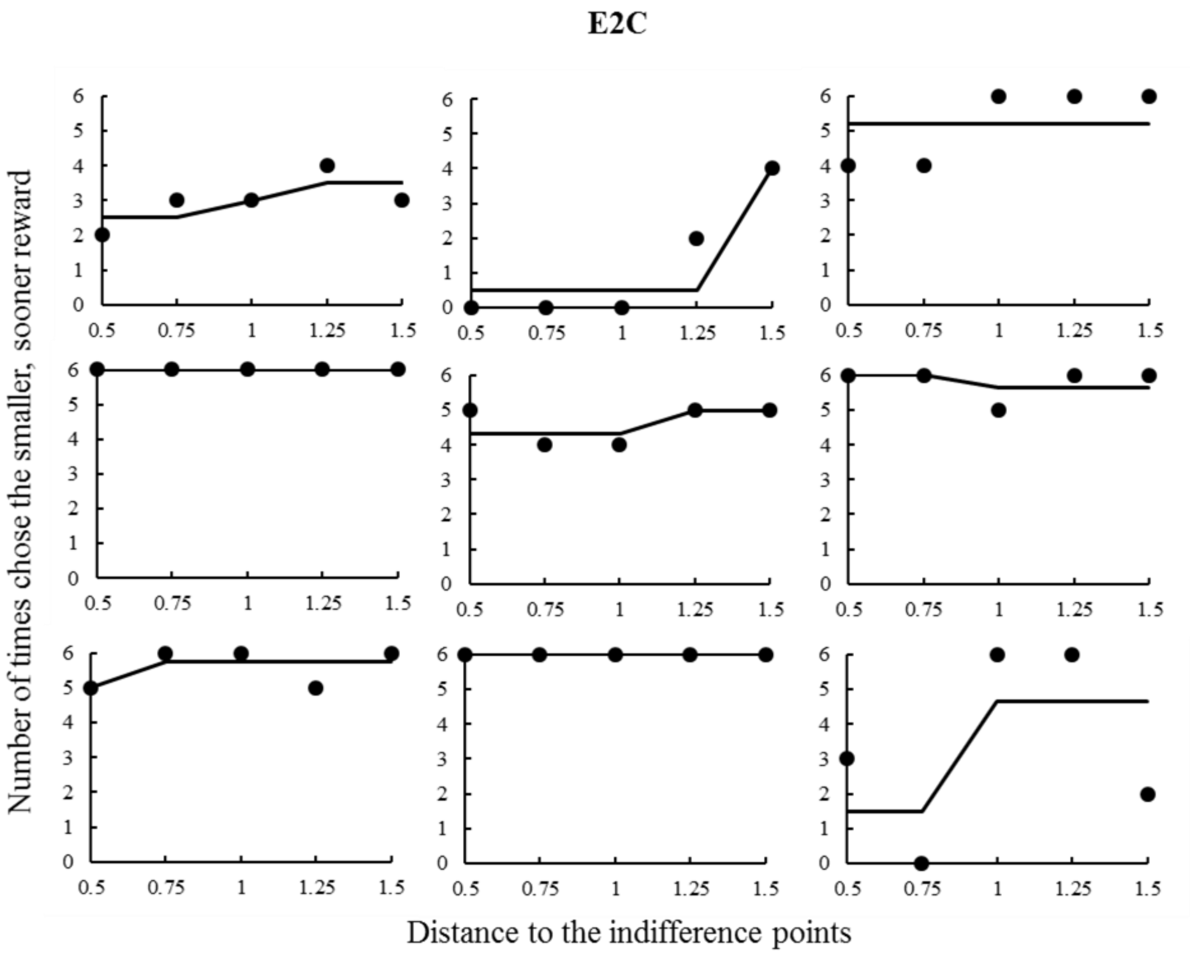

Figure G.3. The logistic function (black line) fitted to the number of times the smaller, sooner reward was chosen at each distance to the indifference point (black points) from Experiment 2C. Individuals who showed unsystematic discounting in the report indifference points procedure and had logistic model R^2s < 0.5 in the distance to indifference points procedure are shown.

Appendix H: Hyperbolic *k* results from Experiment 3A

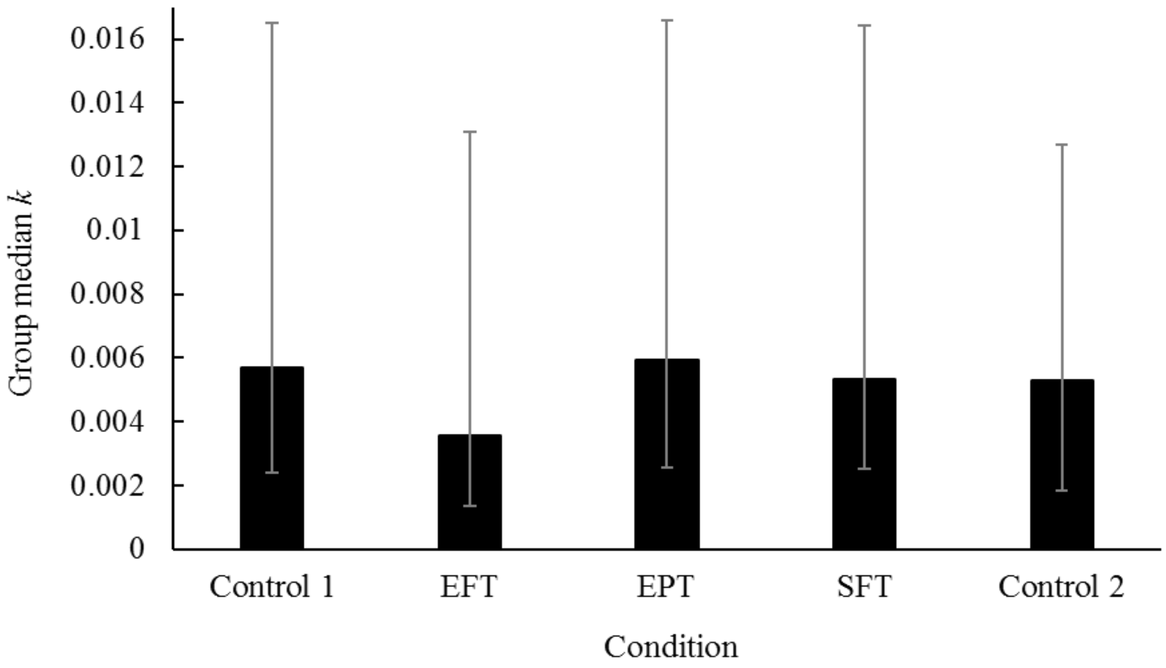

Figure H. Group median *k* values for the control time 1, EFT, EPT, SFT and control time 2 conditions. Error bars represent the interquartile range.

A Friedman test showed a significant main effect of condition on the *k* values, X^2 (4) = 25.63, $p < 0.001$. Table H shows the *p*-values from the follow-up Wilcoxon signed rank tests.

Table H
p values from post hoc comparisons using Wilcoxon signed-rank tests

	EFT	EPT	SFT	Control 2
Control 1	0.008	0.620	0.282	0.023
EFT		0.006	0.025	0.139
EPT			0.244	0.043
SFT				0.594

Note: Bonferroni-corrected alpha level = 0.005.

Appendix I: Plotted residuals for each condition in Experiment 3A (Figure I.1) and Experiment 3B (Figure I.2) based on participants with systematic discounting

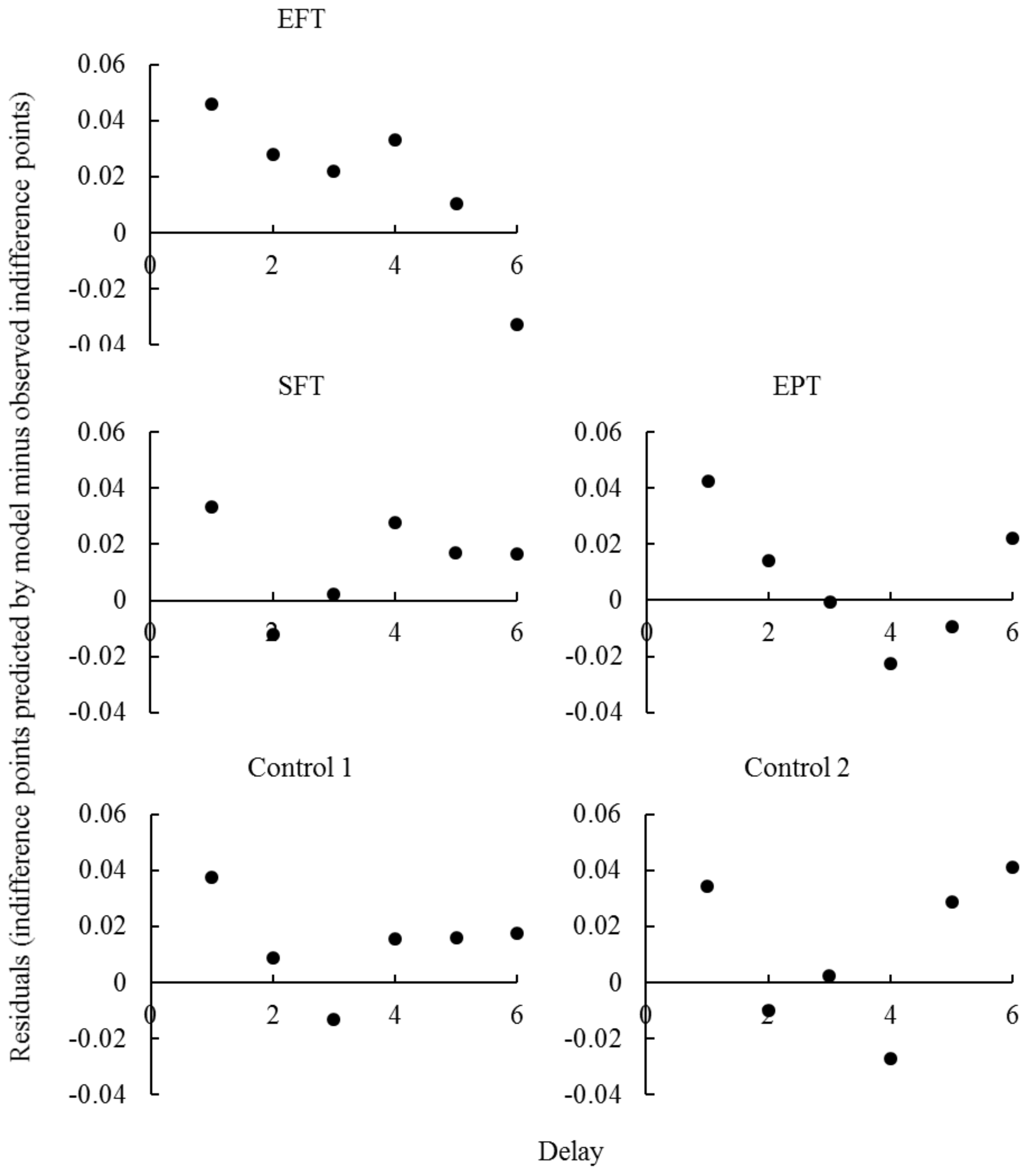

Figure I.1. Median residuals for each delay in the EFT, SFT, EPT, control 1 and control 2 conditions from Experiment 3A based on participants with systematic discounting. Points under the x axes represent underestimation of value by the hyperbolic model whereas points above the x axes represent overestimation of value by the hyperbolic model.

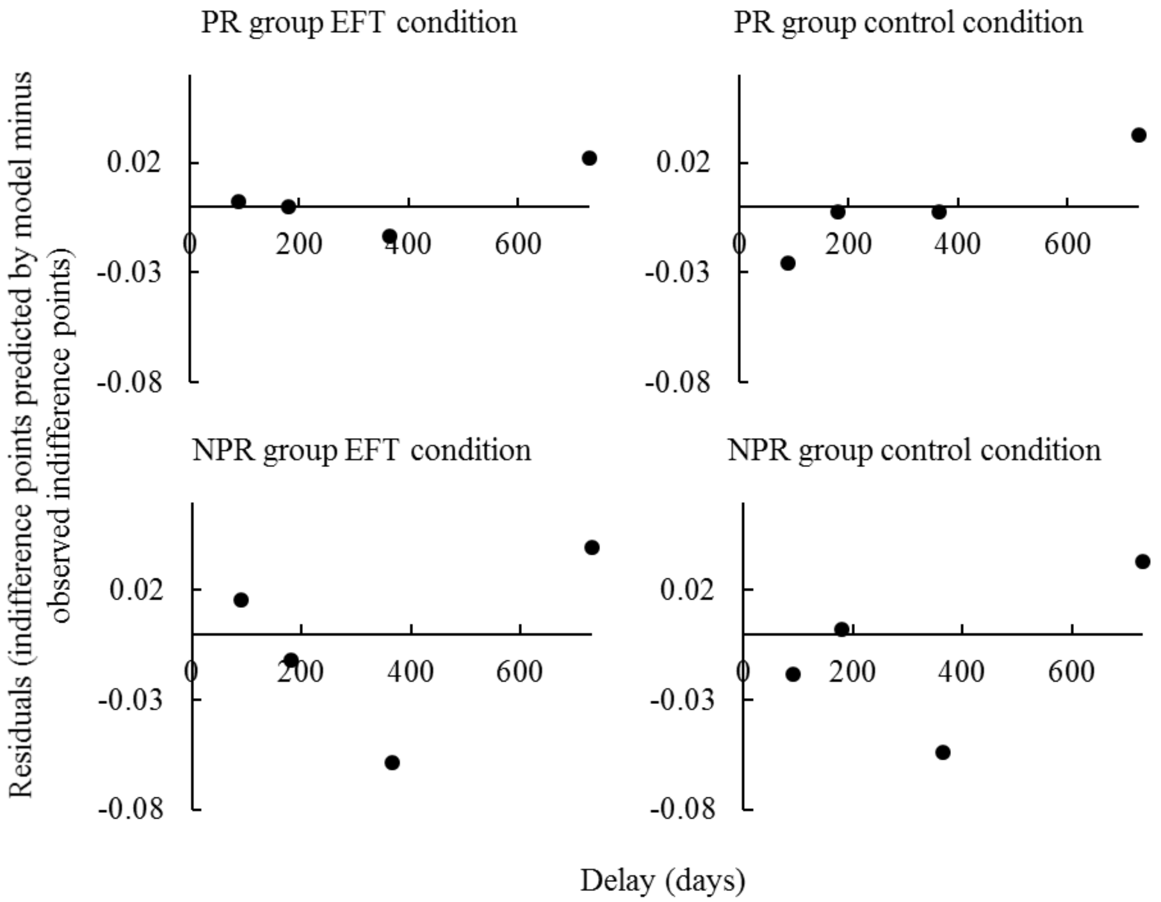

Figure I.2. Median residuals for each delay in the EFT and control conditions for the personally relevant (PR) and non-personally relevant (NPR) groups from Experiment 3B based on participants with systematic discounting. Points under the x axes represent underestimation of value by the hyperbolic model whereas points above the x axes represent overestimation of value by the hyperbolic model.

Appendix J: Friedman test results for each type of event attribute in Experiment 3A

Event attribute	X^2	*df*	*p*
Exciting	23.64	2	0.000*
Expensive	110.12		0.000*
Positive	26.13		0.000*
Imagery score	5.64		0.060

* Statistically significant

Appendix K: List of questions for all events used in Experiment 3B (taken from Yi et al., 2016 & 2017)

Personally relevant EFT condition

Lunch:

1. At what time will you have lunch in [*insert delay*]?
2. What will you have for lunch in [*insert delay*]?
3. Where will you have lunch in [*insert delay*]?
4. What will you drink at lunch in [*insert delay*]?
5. Who will you have lunch with in [*insert delay*]?
6. Enter a cue word/phrase for this activity.

Website:

1. Name a website you will visit in [*insert delay*]?
2. Why will you visit this website in [*insert delay*]?
3. At what time of day will you visit this website in [*insert delay*]?
4. What will be the primary purpose of this website that you will visit in [*insert delay*]?
5. Where will you be when you visit this website in [*insert delay*]?
6. Enter a cue word/phrase for this activity.

Leisurely activity:

1. Very briefly describe a leisurely activity that you will participate in [*insert delay*]?
2. What will you wear for this activity in [*insert delay*]?
3. Who else will be involved in this activity in [*insert delay*]?
4. Where will this activity take place in [*insert delay*]?
5. At what time will this activity take place in [*insert delay*]?
6. Enter a cue word/phrase for this activity

Conversation:

1. With whom will you have a conversation in [*insert delay*]?
2. Where will this conversation take place in [*insert delay*]?
3. What will the main topic of this conversation be in [*insert delay*]?
4. At what time of day will this conversation take place in [*insert delay*]?

5. Why will this conversation take place in [*insert delay*]?
6. Enter a cue word/phrase for this activity

Non-personally relevant EFT condition

The questions for the non-personally relevant condition were identical to those in the personally relevant condition, except that the word ‘you’ was replaced with the person’s name that the participant had previously entered. Additionally, when participants were asked to state who was present for each event, they were reminded that they themselves will not be present or involved in any way. They were also told that the main topic of conversation will not be about themselves when asked about the conversation event.